DOCUMENT RESUME

ED 060 899 LI 003 567

TITLE	Project Intrex. Semiannual Activity Report, 15 September 1971 - 15 March 1972.
INSTITUTION	Massachusetts Inst. of Tech., Cambridge.
SPONS AGENCY	Carnegie Corp. of New York, N.Y.; Council on Library Resources, Inc., Washington, D.C.; National Science Foundation, Washington, D.C.; Office of Education (DHEW), Washington, D.C.
REPORT NO	Intrex-PR-13
PUB DATE	15 Mar 72
NOTE	111p.;(46 References)
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Computers; Economics; *Electronic Data Processing; Information Networks; *Information Retrieval; Information Storage; *Information Systems; *Library Automation; *Use Studies
IDENTIFIERS	Computer Software; *Project Intrex

ABSTRACT

Heavy emphasis was placed on experiments, and interpretation of experimental results. A set of experiments was designed to yield quantitative information on how the experimental subjects used the full-text-access system, why they used it and how effective it was. A detailed report of work on this topic to date is presented. The in-depth analysis of the Intrex system of bibliographic storage and retrieval is continued. The economic studies of information systems were extended along lines that refined the system models being used for study and that included consideration of networks of information systems. Two Project Intrex-designed display terminals are now in operation and both can engage the Intrex system simultaneously. The terminal has been newly named BRISC (Buffered Remote Interactive Search Console). Users prefer the BRISC to other available terminals because of its large-size characters, bright display and the save-page feature of the terminals. Refinements in the full-text-access system have been made to overcome occasional difficulties experienced in centering text on the cathode-ray-tube screen. (Author/NH)

ED 060899

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT INTREX

SEMIANNUAL ACTIVITY REPORT

15 September 1971 - 15 March 1972

Intrex PR-13

15 March 1972

LI 003 567

CAMBRIDGE MASSACHUSETTS

ACKNOWLEDGMENTS

The research reported in this document was made possible through the support extended the Massachusetts Institute of Technology, Project Intrex, under grants from the Carnegie Corporation, the Council on Library Resources, Inc., the National Science Foundation, and the U. S. Office of Education.

TABLE OF CONTENTS

I.	INTRODUCTION	page	1
II.	RESEARCH AND DEVELOPMENT ACTIVITIES (Electronic Systems Laboratory)		3
	A. STATUS OF PROGRAM		3
	B. SYSTEM USAGE: EXPERIMENTS AND ANALYSIS		5
	Summary		5
	Intrex Facilities in Open Environments		5
	Text-Access Experiments		8
	The IAP-Intrex Experiment		16
	The Intrex-Rutgers Experiment		19
	Retrieval Devices and Intrex Subject Indexing		20
	Retrieval Effectiveness, Indexing, and Strategy		21
	Retrieval Effectiveness, Coordination Level, and Search Exhaustivity		29
	Catalog Indicativity Experiments		41
	C. ECONOMIC ANALYSIS		48
	Summary		48
	Modeling of Information-Retrieval Systems		48
	Dedicated Information-Retrieval Design		52
	Information-Retrieval Networks		56
	Network Cost Optimization		62
	D. AUGMENTED-CATALOG INPUTTING		64
	Summary		64
	Processing of Intrex Documents		64
	IAP Data Base		66
	E. COMPUTER SOFTWARE		70
	Summary		70
	Intrex Retrieval-System Software		70
	IAP Retrieval System		72
	Buffer/Controller Software		72

TABLE OF CONTENTS (Continued)

	F. HARDWARE	page	76
	Summary		76
	The Intrex Display Consoles		76
	Full-Text Storage and Retrieval		77
III.	MODEL LIBRARY PROJECT		79
	A. STATUS OF THE PROJECT		79
	B. POINT-OF-USE INSTRUCTION		81
	C. PATHFINDERS		87
	D. USER PREFERENCE STUDY		95
	E. NEW PROGRAMS		100
	F. VISITOR'S PROGRAM		102
IV.	PROJECT INTREX STAFF		103
V.	CURRENT PUBLICATIONS		104
VI.	PAST PUBLICATIONS		104

PROJECT INTREX

Activity Report

I. INTRODUCTION

The applications of computers to libraries and information systems have been the subject of an important new study conducted under the aegis of the National Research Council. Entitled "Libraries and Information Technology — A National System Challenge," it is addressed to the Council on Library Resources, Inc. It was carried out by the Information Systems Panel of the Computer Science and Engineering Board of the National Academy of Sciences. The Chairman of the Panel was Dr. Ronald L. Wigington, Director of Research and Development, Chemical Abstracts Service.

The unique quality of the Wigington Report is its combination of technological competence with organizational wisdom. Information science has reached a new level of maturity when six distinguished panelists from the world of libraries, computers, and information systems agree that

> "The primary bar to development of national computer-based library and information systems is no longer basically a technology-feasibility problem. Rather it is the combination of complex institutional and organizational human-related problems and the inadequate economic/value system associated with these activities. National leadership to solve these problems has not emerged."

We may hope that the National Commission on Libraries and Information Science will soon provide that leadership. In the meantime, it is the responsibility of the information science community to organize its capabilities toward that day. After surveying the state and trends of the relevant technologies, the Wigington Report recommends that

> "The present collection of localized and fragmented efforts must be guided toward harmonious integration through experience with a comprehensive pilot system."

and

> "To develop information systems consistent with geographic dispersion of information resources and information users, increased stress must be placed on scientific design and modeling studies of broadly based information networks."

It is in exactly these directions that the future of Project Intrex is being planned beyond mid-1972, when the initial research stage of our program will reach its conclusion. We are proposing to utilize our research findings in a prototype operational system, regional in character, centered on M.I.T. Its basic purpose will be to organize community involvement with new forms of library operation, and to establish the economic viability of such operations in the university environment.

By scaling up the experimental online interactive system developed by Intrex to prototype operational size and expanding the pattern of services offered to the user community, we expect to find out whether user charges will be acceptable at a level commensurate with actual costs. Network operation will be essential to the realization of that goal.

* * * * * *

With the appointment of Charles H. Stevens as Executive Director of the National Commission on Libraries and Information Science, Project Intrex has lost one of its earliest and most effective leaders. The library orientation of Project Intrex has been his continuing concern, from the days of the Intrex Planning Conference to the present. In all decisions of experimental design that related to the ultimate use of the system in the library, he was the conscience of Project Intrex. He created the Model Library Program to provide those aids to users that will be needed in all libraries that combine conventional and innovative services. The Barker Engineering Library at M.I.T., in its operational concepts as well as its physical form, is the result of years of dedicated effort and unremitting care by Charles Stevens. None of us who have worked with him will forget his loyalty and friendship.

Carl F. J. Overhage
Cambridge, Massachusetts
15 March 1972

II. RESEARCH AND DEVELOPMENT ACTIVITIES (Electronics Systems Laboratory)

A. STATUS OF THE PROGRAM

Professor J. F. Reintjes

Heavy emphasis was placed on experiments, and interpretation of experimental results during the past six months.

A major effort was made to determine the ways in which users employ the full-text-access feature of the Intrex system. A set of experiments was designed to yield quantitative information on how the experimental subjects used the text-access system, why they used it and how effective it was. A detailed report of our work on this topic to date is presented in this issue of the report.

Our in-depth analysis of the Intrex system of bibliographic storage and retrieval continues. The catalog-indicativity experiments described in previous Activities Reports have been expanded to include additional experimental subjects and to provide new data on the retrieval effectiveness of Intrex, as compared with the retrieval effectiveness of indexing services.

In an effort to demonstrate the flexibility of the Intrex information storage and retrieval system and to illustrate a kind of supplementary information service libraries might render through machine aids, we provided a special online information service during MIT's Independent Activities Period, a one-month period of on-campus independent study between fall and spring semesters. A data base was developed on all IAP activities offered by the Electrical Engineering Department, including items such as mini-courses, lectures, seminars, research opportunities, and so forth. One important observation that can be made as a result of this exercise is: supplementary data bases of this kind are easily infused into the Intrex system and computer-software changes can be quickly made to accommodate the new data base. An equally important observation is that an information service of this kind is enthusiastically received by students and others who use it.

Our economic studies of information systems were extended along lines that refined the system models being used for study and that included consideration of networks of information systems.

We now have two Project Intrex-designed display terminals in operation and both can engage the Intrex system simultaneously. The terminal has been

newly named BRISC (for Buffered Remote Interactive Search Console); and it is in brisk demand in the Barker Engineering Library. Users prefer the BRISC to the other terminals available to them because of its large-size characters, bright display and the save-page feature of the terminals. In order to permit two BRISC's to operate simultaneously, changes were required in the buffer/controller software. An upgrading in the software system to simplify BRISC operation was made while the necessary changes were being implemented.

Refinements in the full-text-access system have also been made in order to overcome occasional difficulties we have been experiencing in centering text on the cathode-ray-tube screen.

B. SYSTEM USAGE: EXPERIMENTS AND ANALYSIS

Staff Members

Mr. A. R. Benenfeld
Mr. L. E. Bergmann
Ms. S. F. Brown
Ms. M. A. Jackson
Mr. P. Kugel
Mr. R. S. Marcus
Ms. V. A. Miethe

Professor J. F. Reintjes
Mr. J. R. Sandison

Undergraduate Student

Mr. D. J. Bottaro

SUMMARY

Use of Intrex facilities in the open enviroment has been further studied with special emphasis on the comparative utility of the different computer consoles. Experiments designed to test the utility of the text-access facility have been run on 11 experimental subjects. Approximately 250 records describing planned activities of the M.I.T. Electrical Engineering Department for the January, 1972 M.I.T. Independent Activities Period were added to the data base in order to test the desirability of an Intrex-IAP information service. An experiment in which students of the Rutgers University Graduate School of Library Service access the Intrex system from the Rutgers location has been planned. Additional analysis relating retrieval effectiveness, indexing, and search strategy has been carried out and a presentation on this subject was made at the 1971 annual meeting of the American Society for Information Science. The new series of catalog-indicativity experiments has been extended. As a by-product of these experiments, preliminary results have been obtained about user preferences among the various fields in the Intrex augmented catalog.

INTREX FACILITIES IN OPEN ENVIRONMENTS

General. The Intrex retrieval system has now served over 1,000 serious users, where the term "serious users" is intended to exclude system personnel, demonstrations, and the like. We estimate that the average number of times a given user has engaged the system is two.

During this reporting period, the Barker Engineering Library station has been maintained on a regular basis, and was available to users two hours a day (1-3 p.m.), five days a week and at additional times on request. A typewriter console has also been available at Harvard University. The station in the Bush

Building, where many of the offices, laboratories, and classrooms devoted to Materials Science are located, has been available by appointment.

During this half-year period alone, a total of 303 serious uses and 58 "learning" uses were made of the system. Through the devices of reduced scheduling and limiting access to the more serious users, we have purposely reduced the number of users from its previous high level to optimize utilization of staff effort toward experimentation and analysis of the more important system usages.

Many of our users during this period have used the Intrex retrieval system in conjunction with traditional library facilities. Many of our serious users during this period appear to have been directed to the Intrex system by references in the library's card catalog. Conversely, many of our users have found references through Intrex that led them to the regular library facilities. These references frequently are citations in the text of data-base articles.

Comparison of Usage at the Different Consoles. Since September 27, 1971, the BRISC has been available at the Engineering Library station on a two-hours a day basis. Little difficulty has been encountered in maintaining this schedule. The equipment has been quite dependable and failures to maintain this schedule have been primarily attributable to the staff's desire to take the system down for changes.

In general users say they prefer the BRISC to the ARDS and the ARDS to the DATEL (typewriter) although there are individual users who favor each console over the others. For example, when both the ARDS and BRISC were available, but not the DATEL, at least four users complained about their inability to use the DATEL for Intrex retrieval during the first week it was unavailable.

The Intrex advisers and regular users especially prefer the BRISC, whereas new users are quite willing to start on any console that is offered to them. The more experienced users seem to have learned the advantages of this console. The main reasons offered for preferring the BRISC were: the larger, clearer characters (particularly for the advisers who give demonstrations); the ability to save displayed information; and the ability to leave a page on the text-access screen while continuing to search the catalog. This latter capability is particularly favored by students who use the text-access screen to store the page of a document containing the bibliography and then do author searches on the names of authors cited in the bibliography. The main complaint about the BRISC has been that a full screen displays too few characters or too little information. It is

felt that at least part of the reason for this complaint is the relatively "open" format we have used for displaying information on the BRISC; a more compact format would add information to the display, probably without significant degradation of clarity.

During a fourteen-week period beginning September 27, 1971, the three Engineering Library consoles accounted for the following average amounts of computer (CPU) time per week: BRISC console, 0.52 hour; ARDS console, 0.44 hour; DATEL, 0.12 hour. It should be noted that the amount of time users spend at the console is greater than the computer time by a factor of 10 to 15, the higher factor being applicable to the DATEL because of its slower output rate. Another consideration was the newness of the BRISC; a clearcut learning, or adaptation, curve seems to apply to usage of this console. The first week it was in operation it accounted for approximately half as much CPU time as the ARDS, whereas during the last three weeks of the fourteen-week period, it accounted for more than 1.5 times as much CPU time as did the ARDS. Even this latter figure should be corrected in favor of the BRISC for the fact that users who come to the system while the BRISC is in use often use the ARDS rather than wait, even when they might have preferred to use the BRISC. These figures tend to confirm the preference ordering given above, which was derived from expressed user and adviser opinions.

In the same fourteen-week period, use of the ARDS console in the Bush building averaged 0.10 hour of computer time a week while the Harvard IBM 2741 typewriter console averaged 0.03 hour of computer time. It appears that the relatively smaller use of the Bush console compared to that of the Engineering Library consoles is due primarily to the requirement for an appointment. We have observed that very few users at the Engineering Library make appointments (although this procedure is advisable to avoid conflicts) and that users in general seem to have a great reluctance to make appointments; they prefer to take their chances that a console is available. In previous reports we have explained the low usage at the Harvard console on the basis of its relatively meager operational facilities compared to the other consoles.

Adviser Training. During this reporting period, one additional member of the Barker Engineering Library staff received training as an adviser in the use of the Intrex retrieval system. Time commitments precluded running the formal training and practice program detailed in the Semiannual Activity Report of 15 September 1971. Consequently, the new trainee learned most of the details of

system use from previously trained advisers, after which, the laboratory staff provided four hours of discussion on underlying concepts of the retrieval system.

TEXT-ACCESS EXPERIMENTS

Summary. The purpose of the text-access experiments is to determine the role played by the rapid availability of full text in the use of the Intrex retrieval system and to evaluate it in this role. A procedure for conducting such experiments has been designed and tested. Thirteen experiments, involving eleven users, have been performed. Although such a sample may be too small to provide a satisfactory basis for thorough conclusions, some preliminary observations, at least, are warranted on the basis of the evidence obtained. These observations, which are described in more detail later in this section, can be summarized as follows:

1. Over 80 percent of the sessions involved use of the text-access subsystem. Most users said they found rapid access to full text a crucial element in a fully satisfactory system and that the text-access system, as implemented, was more than fast enough for them.

2. The system operated reliably; over 90 percent of the commands initiated by the user resulted in the specified output.

3. Although most users seem satisfied by the quality of the image when text access is used for the purpose for which it was originally intended (the preliminary examination of the document), most users prefer higher resolution and express the desire to eventually obtain hard copy. Also, approximately half of the viewing time was spent with a magnified image -- a feature which has been included to overcome marginal resolution.

4. Users in this experiment employed the text-access system primarily to judge the relevance of documents after preliminary judgements have been made on the basis of the catalog data. Users employed text access only to a small extent to read document text for information it contained; we identified no explicit instance of text use for search-strategy formulation.

5. Relevance judgements with the full-text system are made primarily on the basis of document text, as such, rather than associated parts such as illustrations, or the abstract. If we assume that academic level (by year of study) is a reasonable measure of a student's depth of knowledge about the subject of his search, then this depth of knowledge

is negatively correlated with the utility of full text for the relevance judging function. The more a user knows about a subject the more he is willing to rely solely on catalog information.

6. The average user spends somewhat more time looking at catalog data than he does looking at text, but since he looks at catalog data on more documents, he spends more time, per document, looking at text.

Design of Experiments. The data for the text-access experiments comes from careful observations of actual user sessions with the Intrex system in the open environment. On a day chosen to run the experiment, subjects are selected for inclusion in the experiment in a basically random manner — as they come to the system with a genuine problem.

Observations are recorded by the computer's monitoring system and by a human observer who notes the behavior that the computer system does not capture. This latter category includes virtually all the user's interactions with the text-access system. In addition to the observer, an adviser is present to assist the user, as in a normal session. Interviews, both before and after the session, provide additional information about the user's background, the nature of his problem, and his evaluation of the system's performance.

Three kinds of data are sought during the session:

1. Descriptive data that characterize, as precisely as possible, how the text-access system is used and how it performs.

2. Motivational data that characterize why the user is doing what he is doing.

3. Evaluative data that characterize how well the system is doing, what the user wants (or expects) it to do, and how well it satisfies his objectives.

A user session is divided into three parts:

1. A pre-session interview during which the suitability of the user as an experimental subject is determined. (Does he have a real problem or is he just interested in trying out, or learning about, the system?) During this phase, the user is given standard information about the system to assure that his behavior is not the result of ignorance. The user's background and the nature of his problem, as he sees it before engaging the system, are also ascertained.

2. The session itself, during which the user is permitted to pursue his problem in whatever way he wishes. The observer

and adviser provide information only when asked, although they will occasionally offer advice when minor technical details seem to be interfering with progress, as for example, when the user forgets to press the carriage-return key.

3. The post-session interview, during which the user is asked to explain features of his behavior when this is needed to give a full account of what happened. During this interview, the user is also asked, by means of a prepared list of questions, to describe and evaluate the results he has obtained, and to express his opinions about system features and performance.

We find that it is desirable to have an observer present in addition to the adviser who is available to assist the user. The data-gathering task requires all of the observer's time while the full-text system is being engaged.

Results. The eleven users who have participated in the experiments to date represent a good cross section of the student population, at least by academic level. There was one freshman and two from each of the three other undergraduate years, three graduate students, and one post-graduate student. All the graduate students and four of the seven undergraduates were previous users of Intrex. In addition to the fact that returning users are a pleasant testimonial to the usefulness of the system, the returning user is a better subject for our purposes because he spends less time than the new user trying things out and learning about the system. Virtually all our experimental subjects had had some previous experience with computers. Other data are summarized in Table IIB-1.

We find that the average console session used 46 minutes of real (console) time and 3 minutes of computer time. Graduate students used more time than undergraduates. The typical computer-time to real-time ratio is 15:1. Note that this figure includes the time spent looking at text during which virtually no computer time is used and this largely explains why the ratio is somewhat higher than the figures observed previously for display console use when the text system was little used.

The average user spends 10 minutes looking at text, or engaging the text system to ask for text. This contrasts to 14 minutes (average) spent with the information from the catalog fields on the screen. Thus, about 42 percent of the total time spent looking at information about documents is spent on the text itself. The average user, however, looks only at slightly more than 10 pages of text from an average of slightly over 4 documents. Since he obtains some catalog

Table IIB-1

Text-Access Experiment Result Summary

Total number of sessions	13
Total number of users	11
Average console real time per session	46 minutes
Average CPU time per session	3 minutes
Ratio	15 : 1
Division of user's real time per session:	
Catalog-field time	14 minutes
Text time	10 minutes
Other time (e.g., search)	22 minutes
Division of catalog-field time (by fields):	
Normal (title, author, journal location)	43 percent
Title only	17 percent
Abstract	19 percent
Match	7 percent
Fiche	7 percent
Division of text time, by function:	
To seek revisions to search formulation	0
To check relevance of document	78 percent
To obtain information from document	12 percent
Division of text time, by kinds of information examined:	
Text only	76 percent
Illustrations	14 percent
Abstract	6.5 percent
Bibliography	3.5 percent
Division of text time (by display mode):	
Unmagnified image	54 percent
Magnified image	46 percent

Table IIB-1 (Cont'd)

Availability of full text of documents in system	75 percent
Reliability of system (percent of requests that succeed):	
Documents obtained	91 percent
Pages obtained	97 percent
Magnifies obtained	96 percent
Search-phase effectiveness (documents per session):	
Number of documents found in session by searches	400
Number of documents for which some catalog information was examined	45
Number of documents for which at least two catalog output requests were made	8
Number of documents for which text was requested from the text system	4.5
Number of documents for which hard copy was obtained	2

Answer to the question "How much delay in accessing the text would you be willing to accept?"

Facility:	Delay which would cause:	
	Minor inconvenience	Major inconvenience
Access to cathode-ray-tube copy of text	15 minute wait	One day wait
Access to hard copy of text	One day wait	One week wait

(above are answers of the median user)

information about 45 documents and makes more than one output request for approximately 8 documents, he spends more time per document looking at text than he does looking at catalog information.

One of the main questions we have asked in the text-access experiments is: "What do users actually use the text-access system for?" In order to answer this question, it is useful to think of a search for information in a retrieval system as consisting of three phases:

Phase 1: A search phase, during which the facilities of the system are used to select a set of pertinent documents. In the Intrex system, such a search is conducted by the computer in response to a search request such as: subject corrosion of indium.

Phase 2: A selection phase, during which the user examines information about the document, or the document itself, to select from the results of his search, the documents that meet his needs sufficiently well to merit reading. Such selection may be made on various grounds, including relevance (of subject matter), recency, length, style, and the like. Much of the information that a user needs during this phase can be found in both the Intrex catalog system and in the text itself.

Phase 3: A reading, or information-obtaining phase, during which the user obtains the information he wishes.

It is not always possible to place a user's activity solely into one of these phases and not every engagement with the system goes through all three phases. However, there are relatively objective ways that the use of the Intrex system can be divided into these three phases. In particular, we judge the user's phase in terms of his commands — for example, he moves to Phase 2 when he types an output command. On this basis, we estimate, using the experimental data, that the user spends slightly more of his time in the search phase than he does in the selection phase. The information that the user obtains from the text images, as provided by the text-access system, seems to be used almost wholly to help with the selection phase, although it might, in principle, contribute to both, or either, of the other phases. Thus, we find that 78 percent of the time that the user spends with the text-access system is used for document selection. Only 12 percent is devoted to obtaining information from the document and, at least in this rather small sample, none is used (as best we can determine) to formulate search strategies.

Interviews strongly suggest that the small percentage of time spent with the text for the purposes of reading the document is the result, not of the marginal quality of the image, but because users want and are used to getting copies of documents to take away with them. These copies are often carried around in a notebook and the user feels free to mark up these copies and to make notes in the margins. Users do not wish to take full notes in absence of such copies. Users in our academic environment also want text copies for later reference since they want to examine details of the text in depth.

The fact that users in this experiment failed to employ information obtained from text images for reformulation of their searches is somewhat disappointing. We anticipated that information obtained from the text would suggest alternative subject-search formulations; indeed, many of the users in the Class experiment (described in the Semiannual Activity Reports for 15 March 1971 and 15 September 1971) did so. We also expected users to search the bibliographies of documents in order to find alternative title and author search formulations, as we had observed other users doing in the open environment. We hypothesize that such uses of the text-access system will increase as users gain more experience in formulating effective strategies in an interactive system, both through their own efforts and through instructional prodding.

Whether we count all of the time spent with text as selection time, or subtract out the 12 percent consumed in obtaining information from the documents, we find that the full text is used more of the time than any catalog field in the selection phase. The average user spends 10 minutes on text, as against 5 on the normal field, 2 each on the abstract and title fields, and 1 each on the matching subject term field and the "fiche" field — the latter field giving him the information he requires to get hard copies.

The time spent looking at text seems to be devoted primarily to reading the text itself (76 percent) rather than the illustrations (14 percent), the abstract (6.5 percent), or the citations (3.5 percent). It is not unreasonable to conjecture that the amount of time reading the abstract would increase if the abstract were not also included in the catalog and that, conversely, the amount of time spent reading citations would decrease if citations were included in the catalog.

Users expressed satisfaction with the quality of the image most of the time, but they seem to prefer to read the magnified form of that image. They spent more than half (54 percent) of their time with the magnified image in spite

of the fact that the initial presentation for any document gives the first page in unmagnified form.

The text-access system performed reliably. Ninety one percent of user requests for documents resulted not only in the proper document but in the acceptable presentation of the first page of that document. Once the first page of a document had been found, 97 percent of the requests for other pages of that document resulted in acceptable display of the proper page. A similar degree of reliability (96 percent) was found in the responses to requests for magnified images.

We can think of Phases 1 and 2 as operations that reduce the size of the set of documents that users still continue to consider potentially useful. Measured in this way, users seem to be able to achieve a 90-percent reduction simply on the basis of information obtained during Phase 1 — at which time the user only learns how many documents were found. The average user retrieved 400 documents as the result of a search and only looked at catalog information for 45 of them. A further reduction, to an average of 8 documents (approximately 80 percent reduction), was obtained by examining the first information output for these 45 documents. Of these 8, only 4.5 (average) documents were requested from the text system, thus yielding a total reduction, on the basis of all catalog information examined, of a second 90 percent. Users requested hard copies of an average of 2 documents. While on the basis of this reduction measure, text access is used to sift out only about 50 percent of the documents, this contribution is counted as quite significant by our users.

Users were asked to indicate what levels of degradation they would be willing to tolerate in the speed of availability of hard copy, and examination copies, i.e., text images on the CRT screen. The median user said he would find a 15-minute delay for an examination copy a minor (but noticeable) inconvenience and a day's delay unacceptable. (It was assumed that a user could request several copies and have a fifteen-minute delay before he could look at all of them, rather than such a delay for each one looked at.) In contrast, with an examination copy available, most users felt that a day's delay in getting hard copy was only a minor (but noticeable) inconvenience whereas a week's delay was unacceptable. We hypothesize, however, that as users become more accustomed to the quicker response times possible with an Intrex system they will be less willing to accept the longer delays suggested by their interview answers.

THE IAP-INTREX EXPERIMENT

Objectives. M.I.T. offers an Independent Activities Period (IAP) for its students in January. During IAP students are free to engage in various scheduled and nonscheduled activities of their own interest and choosing. For the 1972 IAP, Project Intrex and the Electrical Engineering Department collaborated in providing, as an experiment, information about Electrical Engineering IAP activities to interested persons via online terminals using Intrex retrieval programs. The purposes of the experiment were:

1. To provide an information service;
2. To provide opportunities for students to participate in an information-transfer research program;
3. To determine whether this kind of service would be worthwhile on an Institute-wide basis for 1973; and
4. To test how well the Intrex system was adaptable to data other than standard bibliographic material.

In large measure these purposes have all been met, as described below.

Implementation. Early in the fall, we designed procedures whereby the IAP activities information could be incorporated into the Intrex catalog structure. These procedural adjustments were made in a straightforward manner; they are described in detail in Section D. Thus we concluded that the basic Intrex catalog structure could adapt to this new kind of data — as it had to data on personal bibliographic files and news articles.*

Student assistants used the basic information about IAP activities provided by Electrical Engineering department personnel to prepare IAP catalog records. The students inserted these records into the computer using online editing programs. Regular Intrex programs were then run to produce a data base formatted for retrieval. The Intrex retrieval programs were used basically unchanged except for a modified, streamlined instructional dialog. In addition, a special two-page, one-sheet version of the short instructional guide was prepared to help users. New IAP activity records were added to the data base as new activities were announced. In addition to regular updating procedures, a new program was devised so that minor modifications to the activity records could

---

* See Intrex Semiannual Activities Report, 15 September 1971.

be added quickly to the formatted data base without resort to a full update. This facility was important for the IAP data base, where, for example, changes in schedule data or location for an activity had to be made quickly (see Section E for IAP programming details).

Operational Experience. The IAP-Intrex retrieval system was first made available to the general M.I.T. community during the second week in December from the DATEL typewriter terminal in the Barker Engineering Library. In early January, coinciding with the start of IAP, expanded access to the system was made available from the ARDS console at Barker and from typewriter consoles in the Electrical Engineering department headquarters and the Student Center Library as well as, on occassion, from a mobile DATEL unit set up in the lobby to the main entrance of M.I.T. System usage was largely on a self-service basis; only rarely were advisers available for help.

The IAP-Intrex retrieval system was, in general, enthusiastically received and heavily used. A total of 192 different persons were identified through our monitor files as having used the system in the period from December 9, 1971 through January 31, 1972. These users engaged the system for a total of 312 sessions. The total number of persons who used or were exposed to the system was considerably higher than these figures indicate because many persons used the system without properly identifying themselves. Others observed direct use of the system by their friends or Intrex advisers.

Results and Observations. Some differences between the IAP-Intrex users and the regular Intrex users were observed. Whereas regular Intrex users are almost evenly split between undergraduates and graduate students, the IAP-Intrex users were predominantly undergraduates by a ratio of 2 to 1. This difference is partly due to the fact that IAP is more heavily oriented toward the undergraduates. Conversely, the regular Intrex data base, being centered on the professional literature, may be somewhat more on the level of the graduate student. For both the regular Intrex and the IAP data bases, faculty and staff comprised about 10 percent of the total users. About two-thirds of the identified IAP users were associated with the Electrical Engineering department; the rest were widely scattered over 11 other departments from 6 users in Physics to one in Political Science. This result, of course, derives from the fact that the data base covered only Electrical Engineering activities and, to a certain extent, from the location of the terminals.

IAP-Intrex sessions were considerably shorter than regular Intrex data-base sessions; the average IAP session lasted about 20 minutes and used about one minute of CPU time whereas the average Intrex session is about twice as long in both respects. Evidence was found that some IAP users want the information system to be available around the clock, or at least for very large daily segments. Regular Intrex data base users have expressed similar feelings. When the IAP terminals were left operating and available on a 24-hour basis we found users engaging the system at all hours of the day and night.

Users were able to operate the system rather well on their own, perhaps even better than do regular Intrex data-base users. We can account for this by the simpler nature of the typical IAP question, the simpler nature of the data base and the absence of associated text. These simplifications, in turn, enabled us to simplify the instructional procedures, and so further help new users get started.

As is the case with the regular Intrex system, the great majority of IAP users were highly enthusiastic and favorable toward the system. Again, as with the regular Intrex system, the biggest single problem seemed to be the incompleteness of the data base; most users, and many of those who declined to use the system, commented on the desirability of covering the full range of activities at M.I.T., not just those of the Electrical Engineering Department. In general, the experience this year suggests that there would be a high utility of an institute-wide system of this kind. In the full system we would want closer coordination with the related information services like printed publication of activity abstracts and schedules. In fact, it may be important to provide publication services as a by-product of the computerized data base to establish the economic viability for this kind of system.

The 250 records in the IAP-Intrex data base were maintained as a separate data base for maximum efficiency of computer operation. However, these same records were also added to the regular Intrex data base. Because the subject matter of the two data bases was largely disjoint,there were not many Intrex users who happened to find the IAP information. A sampling among the few users who did have this experience suggests that this kind of information provides potentially valuable additions to a bibliographic information data base, especially in terms of pointing users to current work in subjects of their interest and, particularly, to faculty engaged in such research who could be contacted directly. The occurrence of activity information among standard bibliographic items may be unanticipated by

users. As we have previously observed with the introduction of augmented bibliographic data in the regular Intrex system, unless some effort is made to explain the nature of the novel information, and the uses to which it could be put, users may ignore it.

THE INTREX-RUTGERS EXPERIMENT

An experiment primarily on retrieval effectiveness is currently being designed in which doctoral candidates at the Rutgers University Graduate School of Library Service will participate. The students will access the interactive Intrex retrieval system using a portable communications terminal and an acoustical coupler with ordinary dialed-telephone-line communication between the terminal located on the New Brunswick campus of Rutgers and the M.I.T.-modified 7094 computer (CTSS) located in Cambridge.

Dr. Susan Artandi, Professor of Library Science at Rutgers, has expressed interest in having Rutgers students gain educational experience in remotely accessing and searching the free-vocabulary, interactive Intrex retrieval system. Through further discussions with Dr. Artandi, a program for such participation is being developed which has, in addition to an educational goal, a research goal of gathering and analyzing additional data to increase our understanding of the retrieval effectiveness of the search process.

Present plans are that system demonstrations and related discussions will be given to approximately 50 interested master's level students. Approximately a dozen doctoral candidates will participate in a formal experiment in which about 1.5 hours of console time will be made available to each student. The experiment and attendant demonstrations are expected to take place in early March, 1972.

A necessary prelude to this remote-access experiment was the successful testing of the telephone-line communication link between a portable, modified DATEL 30 communications terminal at Rutgers and the M.I.T. computer system. We determined that it is possible to access the M.I.T. computer from Rutgers on a dial-up basis, and that a conditioned phone line containing special compensatory features is unnecessary. The Intrex system has been exercised continuously from the DATEL terminal at Rutgers for a two-hour period. During this interval, Intrex operation was flawless, and no difference between Rutgers operation and on-campus M.I.T. operation could be observed.

RETRIEVAL DEVICES AND INTREX SUBJECT INDEXING

A presentation on "Intrex Subject Indexing and Its Relation to Classification" was made in Denver at the 1971 annual meeting of the American Society for Information Science, Special Interest Group on Classification Research. The SIG/CR session concerned itself with views of classification. A report enlarging upon that presentation is in preparation; its major points are briefly highlighted below.

The Intrex retrieval system contains a natural-language indexing vocabulary which is manually precoordinated into English nominative constructs, called index expressions, and each expression is assigned a range number that reflects the depth of the indexing. These index expressions are decomposed into individual, stemmed words which are stored in an inverted file. In the inverted file each stem reference retains information about the context in which the stem appeared, including such facts as the document number, the number and range of the expression containing the stem, and the word ending and word position of the stem within the expression.

Users interacting with the retrieval system use their own natural vocabulary to create a search expression which is then decomposed and stemmed by the system. The basic matching algorithm coordinately matches the user's stemmed words with the inverted file stems. The user has available to him, and under his control, several system commands which allow him to control for variations in vocabulary usage, and to make different combinations of retrieved lists. These commands include the ability to override the stemming process and the simple coordination mechanism. We have evidence that online interactive retrieval systems under user control and employing natural vocabulary indexing are more effective than systems with either an authority vocabulary or systems with retrieval controls not in the hands of users. In the interactive situation, each user may readily tailor his own search strategies to reflect his own use and interpretation of vocabulary.

An indexing language, together with the retrieval system operating on that language, may provide certain features to enhance system operation. Our experimental work to-date on retrieval effectiveness suggests that these features can be rank-ordered in importance as follows:

1. Phrase Decomposition (Single-Word Matching)
2. Word Stemming

3. Natural Vocabulary
4. Boolean Combinations of Words
5. Linking (Word-Position Controls)
6. Stemming Controls
7. Controls on Index Expression Ranges (Weights)

In an online interactive environment, the highest initial performance level is achieved,on the average, for most users when the basic search mechanism is the simple coordination of stemmed, natural-vocabulary words. Individual performance can then be improved when the user is able to call upon the other indexing and retrieval features that may be warranted in the particular circumstances of his search problem. Our experimental work shows that phrase decomposition and stemming of search expressions are important operatives in all retrieval systems, even those using a controlled vocabulary, and that the normalizing function of a controlled vocabulary is, at best, elusive. This elusiveness is attributable, at least in part, to the fact that the normalizing function is inconsistently applied in practice. Natural-language vocabulary is the single best precision discriminator, although in some circumstances clumpings such as a gross-level classification can improve upon initial retrieval performance.

RETRIEVAL EFFECTIVENESS, INDEXING, AND STRATEGY

The number of cases for analyzing retrieval effectiveness and search strategies is being expanded beyond the Class Experiment search problems reported in the Semiannual Activity Report of 15 September 1971. Included in the new cases are the search problems of the experimental subjects (ESs) participating in the indicativity series of experiments. Detailed problem statements and relevance judgements are available from that experimental series which make the cases natural candidates for study of retrieval effectiveness.

The methodology employed in the series of analyses on retrieval effectiveness and search strategy is briefly summarized as follows: A set of relevant documents is identified for use as a recall base. For the cases from the indicativity experiments, this set of documents is taken to be those documents which ESs rated highly useful or somewhat useful (ratings 1 or 2) on the basis of the full text. Once the recall base has been identified, the Intrex indexing of each document is compared to that of the other documents in the set, primarily by noting word stems common to more than one document in the set. From this comparative study,

together with a study of the elements of the original word statement of the ES's search problem, an optimum strategy is propounded for maximum effective retrieval of the recall base. The propounded optimum strategy is then used for searching the Intrex data base. The set of newly retrieved documents, or a sample of the set, is submitted to the ES for judgements which serve either to confirm or to refine the propounded optimum search strategy with respect to the larger, full data base. Data concerning the precision of the optimum search strategy is derived; lower values of precision imply greater user effort in terms of reducing the set of retrieval documents to the relevant ones. Simultaneously with the development of an optimum search strategy for retrieval from the Intrex data base, abstracting and indexing services are searched for the documents comprising the initial recall set. The indexing of these documents by each service is compared in a manner analogous to that for Intrex, and strategies applicable to each service developed. In this manner, the retrieval effectiveness of controlled-vocabulary indexing can be compared to the retrieval effectiveness of the uncontrolled, in-depth, natural language Intrex indexing and to "automatic" indexing of text vocabulary. In addition, having derived optimum search strategies for a variety of situations, we shall be in a better position to suggest procedures by which the interactive search process can best be directed toward discovering the good strategies.

At the time of writing this report, we are in the process of completing the retrieval-effectiveness analysis for the first case drawn from the indicativity experiment series, namely, that for the search problem presented by ES 27 on the topic 'delamination'. The major results obtained to date are briefly summarized.

The initial Intrex recall test base contains seven documents. A comparative analysis of Intrex indexing of the seven documents led to the development of a hypothesized optimum compound search strategy for Intrex retrieval utilizing three major themes associated in some way with delamination. The specific themes, their union and their respective Intrex search strategy vocabularies are:

	Theme	Search-Strategy Vocabulary
(a)	delamination	delamination
(b)	fracture of laminates and composite materials	fracture AND (laminar OR laminates OR composite! OR composites!)
(c)	transverse fracture	fracture AND transverse
(d)	a OR b OR c	delamination OR [fracture AND (laminar OR laminates OR composite! OR composites! OR transverse)]

In the search-strategy vocabulary, 'laminates' stems to 'lamin+' and will pick up other stems with endings such as +ate, +ated, +a, but the word 'laminar' is its own stem. The explicit word forms 'composites' and 'composite' (signified by the !) are used to prevent retrieval of documents based only upon their stem, 'compo-sit+', a stem also common with the more frequently appearing word 'composition'.

The recall effectiveness of the compound strategy and its three major components is illustrated in Fig. IIB-1. Cumulative recall is plotted against the

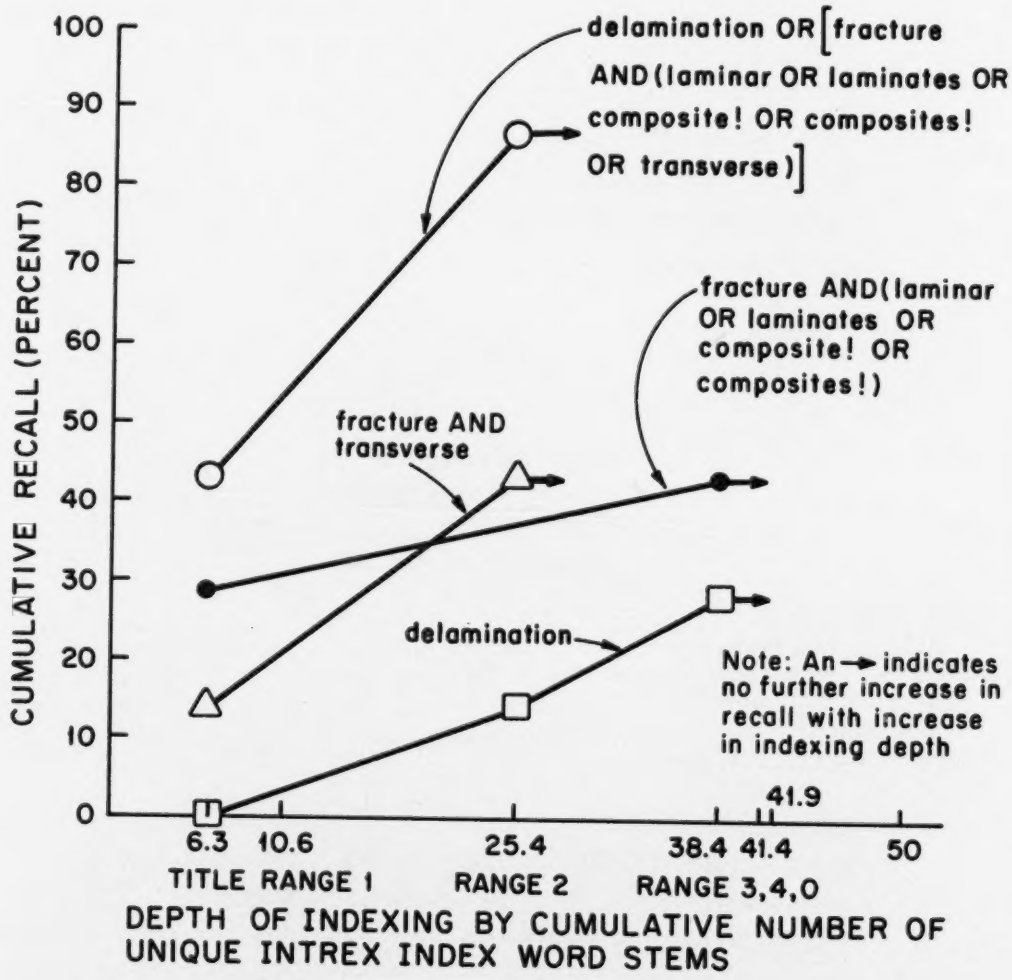


Fig. IIB-1 Intrex recall effectiveness for an optimum search strategy and its components as a function of indexing depth by number of unique word stems cumulated over index range numbers.

cumulative number of unique word stems in the Intrex indexing associated with the recall base, where the cumulation of unique stems is by successively deeper ranges of indexing. The index range number corresponding to each point is shown and the order is title (or range 5), followed by ranges 1, 2, 3, 4 and 0. The combined

strategy is considerably better than any single major component at all depths of indexing; 86 percent recall is achieved at a depth of range two. Range-one index terms did not add to the recall effectiveness of the title words for any of these strategies, nor did ranges four and zero contribute to recall. The data fit the model described in the next section, which explains the relationships between recall, depth of indexing, coordination level, and search term exhaustivity.

It is of interest to compare the optimum strategy achieved on Intrex indexing with the same strategy employed on text. Figure IIB-2 plots recall

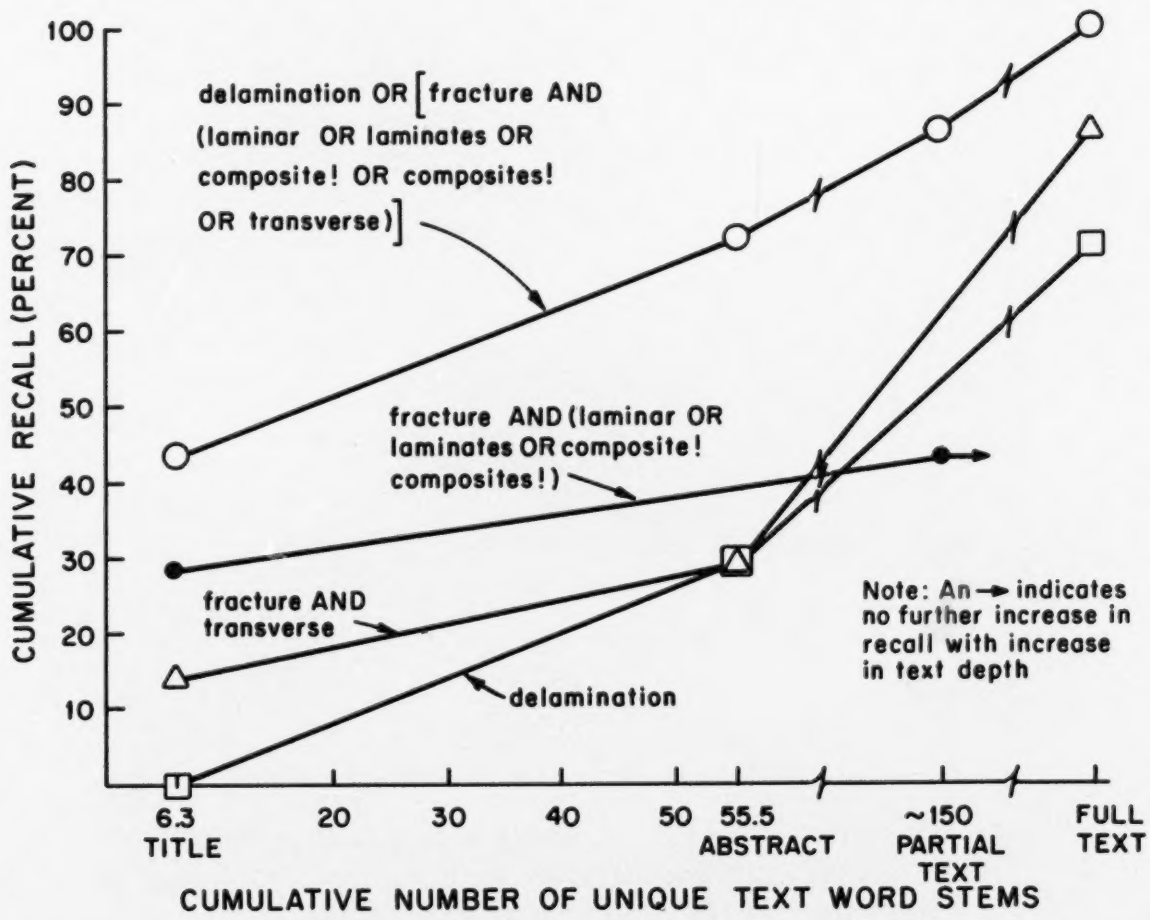


Fig. IIB-2 Recall effectiveness for Intrex search strategies as a function of cumulative number of unique text word stems.

effectiveness versus cumulative number of unique text word stems. For title and abstract the actual number of cumulative unique stems for this seven-document recall base is used; for partial text (consisting of title, abstract, introduction, and conclusion) this number is estimated at about 150 unique stems. No estimate

was made for full-text unique stems. Intrex indexing performs better than the abstract for all strategies, in that an equal or greater recall effectiveness is achieved with a significantly smaller number of unique word stems. Text recall at a comparable level to that attained by the compound Intrex strategy is not achieved until partial text is included, and this is also true for two of the three component Intrex strategies. Complete recall from text is achieved only when full text is used and then only for the compound strategy. Because text vocabulary forms the basis of Intrex indexing, it is not likely that some other search strategy operation on abstract, partial text, or full text, would achieve better retrieval effectiveness in terms of both recall and precision.

Several abstracting and indexing services were searched for the seven documents in the Intrex recall base in order to analyze the functioning of a controlled vocabulary with respect to this search topic. These services were Chemical Abstracts (CA), Engineering Index (EI), Metallurgical Abstracts (MgA), Metals Abstracts (MA), Physics Abstracts (PA), and Review of Metal Literature (RML). Table IIB-2 shows the recall base for each service, and for that base there are shown: the number of access points per document provided by the service exclusive of cross references; the number of unique index word stems per document; the average number of title-word stems; the average number of index-word stems that are and are not also title word stems; and the recall percentage when the strategies developed for Intrex are applied against the index-word stems in these services. In the derivation of the number of unique stems, the complete index line, including any modifiers, was considered, but cross references, if any, to or from the leading word were excluded.

Table IIB-2 shows that the application of only one of the three major Intrex search strategies yielded non-null results; and even that strategy failed completely in two of the six services. These non-null retrievals for the four services yielded only two of the seven documents in the Intrex recall base; indexing based solely on title words would have been sufficient for retrieving one of those two documents with that same strategy. Except for EI and RML, the number of unique index-word stems used by the services appears to equal or slightly exceed the number of unique title-word stems. Generally, however, only 50 to 60 percent of the specific index stems used are also title word stems. This suggests that for purposes of retrieval systems, it would be a significant advantage to add title words to the indexing provided by the abstract service.

TABLE IIB-2

Optimum Intrex Search Strategies Applied to
Abstracting Service Recall Bases

Abstracting and Indexing Services	CA	EI	MgA	MA	PA	RML
Recall Base (Documents)	4	6	2	1	1	3
Number of Index Line Entries per Document	1.0	1.0	1.5	6	2	4.3
Unique Index Line Word Stems per Document	7.3	2.0	4.5	15	4	9.7
Title Word Stems per Document	6.3	6.7	4.5	13	4	5.3
Number of Index Word Stems in Title, per Document	3.8	1.2	2.5	7	4	3.0
Number of Index Word Stems Not in Title, per Document	3.5	0.8	2.0	8	0	6.7
Recall Percent Using Intrex Strategies:						
delamination	0	0	0	0	0	0
fracture AND transverse	0	0	0	0	0	0
fracture AND (laminar OR laminate OR composite! OR composites!)	25	0	100	0	100	67
delamination OR [fracture AND (laminar OR laminate OR composite! OR composites! OR transverse)]	25	0	100	0	100	67

Does some other strategy exist which would optimize the recall results for a given service? For RML, 100 percent recall of three documents could be achieved by a slight broadening of the search to "(fracture OR fatigue) AND (laminar OR laminate OR composite! OR composites!)". However, with respect to MA, (the successor to MgA and RML), this broadened strategy will not retrieve the one document in the MA base, which is not among the documents in MgA or RML. To achieve 100 percent recall of the four different documents found in those three services, the original Intrex strategy must include a new theme: "(crack AND

propagation) OR [fracture AND (laminar OR laminate OR composite! OR composites!)]". For CA, there does not appear to be any reasonable optimum strategy which will recall more than 50 percent of the four documents in the recall base. EI, which has the largest recall base of the services, also has for that base the least depth of indexing with respect to specificity of meaning and number of unique words, and no optimum search strategy is evident.

Cross references to the index terms actually used by the services were not included in any of the index-stem counts. However, in the course of searching the services, it was noted that many cross references are of the "use" type such as 'Brittle Fracture, see Fracturing, and see Brittleness' or 'Lamina, see Laminates', or 'Notch Impact Strength, see Impact Strength'. These types of references invoke phrase decomposition and stemming as user aids. Thus, we would not expect the inclusion of additional words from cross references to significantly add to the indexing picture and retrieval effectiveness values given above,because they are already based upon phrase decomposition and stemming of each index line. It should be noted that a user performing normal, manual searches of these services would have considerable difficulty in carrying out the phrase decomposition and stemming to the extent considered in this analysis. Note from Table IIB-2 that the number of index line entries, which is equal to the number of filing words for those documents, averages only slightly more than one-third the number of unique word stems in the entire index lines for the same documents. Thus, without phrase decomposition, retrieval would be severely inhibited.

Any optimal strategy for retrieval effectiveness must consider precision in addition to recall. Although we have only discussed recall effectiveness up to this point, all the strategies discussed above have qualitatively taken precision into account. With respect to the topic 'delamination', the search word 'delamination', yields the highest precision. Of three documents in the Intrex data base indexed under 'delamination', only one was rejected by the ES, and that for reasons relating to the superficiality of document quality and not for topical irrelevance. We note, in addition, that the word 'delamination' appeared in the full text of five of the seven documents in the recall base, in the abstract of only two documents, and not at all in a title. However, delamination was an index filing word in only one of the six index services, namely in MA and in the form 'delaminating', but it never appeared internally within an index phrase or as a cross reference in any of the six services.

However, the single document in MA that is also in the recall base was not retrievable under 'delamination'.

For reasons of precision, 'delamination' is the only strategy theme embodying a single search word. Each of the other two themes employs an intersection, although the combined strategy is, of necessity, the union of the three major component themes. Although quantitative precision figures for this optimum strategy are yet to be determined through feedback with the ES, the overall list sizes are of some interest. These figures, shown in Table IIB-3, are for the Intrex inverted files for February, 1972. Intersection of search words considerably reduces list sizes. The combined strategy operating on the current Intrex

Table IIB-3

Retrieved List Sizes Using an Optimum Intrex Search Strategy

Search Words	List Size (Documents)
delamin + ation	3
composit + ion	1003
composite!	193
composites!	154
composite! OR composites!	260
fractur + e	450
laminar	24
lamin + ate	43
laminar OR laminate	66
transvers + e	615
(laminar OR laminate OR composite! OR composites!)	307
fracture AND (laminar OR laminate OR composite! OR composites!)	34
fracture AND transverse	18
delamination OR [fracture AND (transverse OR laminar OR laminate OR composite! OR composites!)]	51

data base yields 51 documents. This figure is certainly tolerable in terms of a searcher's effort to judge the utility of that number of retrieved documents.

The single word 'fracture' yields 100 percent recall for the Intrex text base, and also for the MA, MgA, and PA recall bases. By itself, however, it is far from optimum as a search expression in Intrex, or any other data base, because of the excessive sizes of lists it generates. In the Semiannual Activity Report of 15 September 1971, we noted how simple classification, at a fairly gross level, can improve precision. In the present case, fracture is a broad subject area in itself, delamination being only one type of fracture. It appears likely that 'fracture' as an index word functions in the compound optimum strategy in the same manner as would 'fracture' as a class term in a simple-level classification. 'Fracture' is only one level removed from the general modifier "mechanical properties" found in several indexing services. Further studies of classification, as well as other retrieval effectiveness features, are in progress.

Additional discussion of how the results of the analysis in this session fit into the results obtained from other analyses is given in the next session.

RETRIEVAL EFFECTIVENESS, COORDINATION LEVEL, AND SEARCH EXHAUSTIVITY

Formulation of the Problem. In the previous semiannual report we showed evidence that the coordination level of the search strategy affected the relationship between retrieval effectiveness and depth of indexing. We have now found additional evidence to support the previous conclusions and to extend these conclusions into a larger model where other factors, such as level of search exhaustivity, have been included.

These additional developments have emerged from further analysis of the problem of ES 12 which was first reported on in the Semiannual Report of 15 March 1970. The topic of ES 12 was particularly relevant to the questions at hand because it involved fairly high levels of search-term coordination, a sizable number of relevant documents were in the Intrex data base, and a good deal of information on this search was available from other experiments.

The research topic for the doctoral dissertation of ES 12 (The Thermomechanical Processing of Aluminium Alloys) can be considered as the coordination of four broad concepts. These are: (A) the mechanical processing concept, (B) the high-temperature concept, (C) the material concept, and (D) the property concept. A set of terms related to the various concepts is given in Table IIB-4.

Table IIB-4

Terms Used in Retrieval Effectiveness Analysis for Topic of ES 12

	CONCEPT			
Relatedness Level	A Mechanical Processing	B High Temperature	C Material	D Property
1 (U) (User term)	rolling deformation	hot	aluminum alloys $CuAl_2$ (MX)* two-phase (MX) inclusions	hardness yield stress
2 (N) (Near relation)	recrystallization working recovery	high-temperature	[Al or aluminium and (copper or Cu or alloy)] second-phase (MX)	yield strength yield strain (MX)
3 (F) (Far relation)	polygonization fracture microstructure	[(elevated or raised)-temperature] (MX) thermomechanical		ductility mechanical-properties
Number of terms	8	4	6	6

*(MX) indicates term not used as metals abstracts index term.

Besides being tagged by the concept they relate to, these terms are differentiated by their relationship to the user's (ES 12) original search terms. Those terms used by ES 12 in his original search statements to Intrex are tagged with a U. Terms that may be described as near relations to the U terms are tagged N. The near relations include morphological variants of the U terms and other synonymous or closely related terms, especially those appearing in the introduction to the thesis of ES 12. Other related terms that had some bearing on the retrieval of relevant documents of this topic are tagged F for Far relation.

An initial recall base of 8 documents was determined. Documents in the recall base either were rated relevant by ES 12 in previous experiments or were in his bibliography.

Coverage of Search Terms in Several Indexes. In Fig. IIB-3 we present a summary of the extent to which the search terms have been included in the indexing of Intrex and Metals Abstracts (MA) or its predecessor, the Review of Metal Literature. Abstract-word indexing is also considered. For each indexing type, the percentage of search terms actually found in the given index, averaged over documents in the recall base, is given. For Intrex the coverage by a given range, or depth, in indexing is taken to be cumulative over all more important (lower valued) ranges, including titles. Similarly, abstract-word indexing is

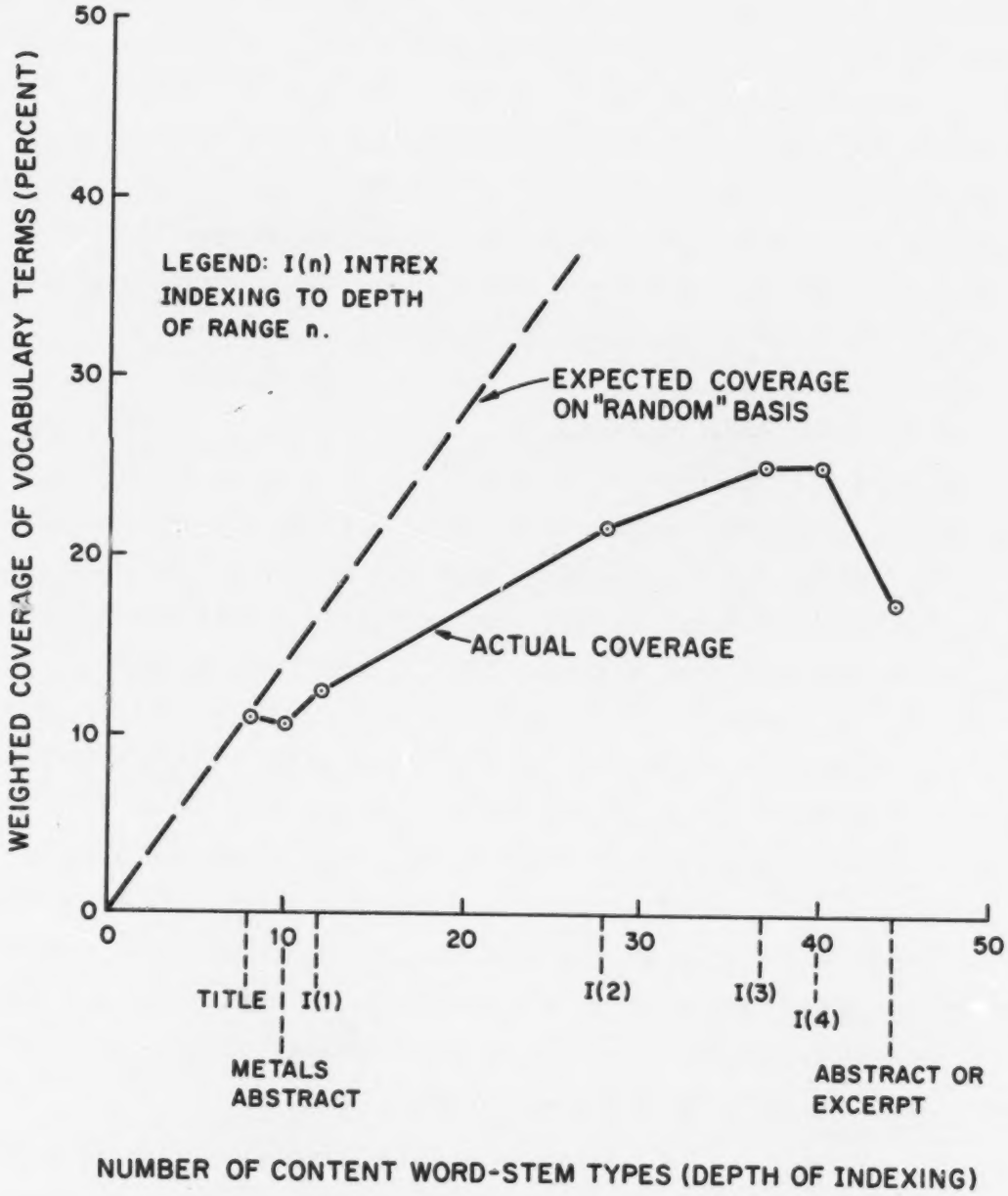


Fig. IIB-3 Weighted Vocabulary Coverage by Index Type

taken to include the title words. The recall base for MA is only 7 documents since one of the 8 documents of the primary recall base was not found in MA. The depth of indexing is measured by the number of content word-stem types. The numbers are estimated averages for the three types of indexing: Intrex, MA, and abstract word.

Several observations on the information of Fig. IIB-3 can be made. Firstly, range 4 and 0 terms add almost nothing to retrieval; and this suggests that the role of tool or technique is probably unimportant for this topic. Secondly, MA indexing does somewhat less well than title-word indexing rather than somewhat better as it has done on some other problems (see, for example, the discussion of the search on "irradiation embrittlement" in the last semiannual report). This result may be attributed primarily to lack of coverage of the "two-phase" concept by MA, whereas no such glaring omission was noted in these other problems. It will also be seen from Table IIB-4 that five of the 24 terms for this topic (21 percent) were not headings in MA; the corresponding percentage for the irradiation embrittlement topic was only 8 percent of 50 terms considered, and the 4 missing terms were relatively unimportant to that search. The failure of the abstracting services to adequately index the delamination topic resulted in even poorer relative retrieval performance, as described in the previous section.

Thirdly, abstracts seem to do relatively poorly on a per-stem-type basis. The most important factor here seems to be that only the 13 common words have been eliminated in the calculations — while this policy eliminates most non-content words from Intrex subject terms, it does not do nearly so well for abstract words (e.g., consider an abstract that begins: 'In this review we consider...').

Fourthly, there appears to be a diminishing-returns effect. In Fig. IIB-3 the slope of the curve from the origin to the title is 1.33 percent/stem-type. The slope from the title to range 2 is 0.56 and from range 2 to range 3 is 0.36. Since the increase in depth is measured in terms of new word stems, on a purely random basis we would expect a constant slope. Thus the deviation from this simple straight line may be attributed to the fact that this sample of 8 from the recall base is definitely not random: it is a collection of articles highly relevant to the topic and so may be expected to have a preponderance of the relevant terms in the important range numbers, as this curve shows.

Recall Analysis. We shall now consider how the indexing coverage of the topic vocabulary affects actual retrieval. The first retrieval parameter we want to investigate is recall where the independent variables are the type and depth

of indexing and search strategy. The particular search strategies considered consist of the disjunction (OR-ing) of one or more terms for a given concept and each concept coordinated (AND-ed) to a maximum of all 3 other concepts. The terms are just those 24 from Table IIB-4.

Since there are 4 levels of coordination and 3 levels of word relatedness, there are 12 search strategies. The recall results for 11 of these strategies are shown in Fig. IIB-4. The 12th strategy, designated U4 for all 4 concepts indexed at user-term level, is not shown explicity — no document matched for any index scheme: recall = 0. Some additional explanation of the meaning of the different strategies may be helpful. Let A1, B1, C1 and D1 be the searches resulting for a simple disjunction of first-level terms in the different concepts. Thus A1 = subject (rolling OR deformation), B1 = subject hot, and so forth. Then U3 results from (A1 AND B1 AND C1) OR (A1 AND B1 AND D1) OR (A1 AND C1 AND D1) OR (B1 AND C1 AND D1). Similarly, if level two disjunctions are indicated by A2, B2, and so forth, (where, e.g., B2 = subject hot or high temperature), then N1 = A2 OR B2 OR C2 OR D2. In order to reduce the complexity of the figure, the Intrex index levels of range-1 terms and all terms have not been explicitly indicated. Range-3 indexing gives exactly the same results as for all terms and, similarly, range 1 is equivalent to title-word indexing in terms of search results for this topic and the given strategies.

The results provide confirmation of the results in the section on coverage of search terms. Firstly, MA indexing is slightly less good than title-word indexing: for five strategies titles are better, for four, MA is better and for three, they are the same. Secondly, abstracts are considerably worse than complete Intrex indexing: 9 strategies favor Intrex, one favors abstract words, and two have the same results. Abstracts appear approximately as effective as range-2 Intrex indexing: 6 strategies favor range 2, 4 favor abstracts, and 2 have identical results.

The important additional information provided by these results is a strong confirmation of the hypothesis concerning the effect of coordination level. Namely, as the coordination level increases, there is a greater chance that the law of diminishing returns is contravened and that there will be a proportionate <u>increase</u> in recall as depth increases (as with strategy U2) or an even greater than linear increase (as with strategies U3, N4, and F4). (For F4 — as well as any other strategy — one may ignore the MA point as not representing a point on a continuous spectrum for a <u>single</u> type of indexing: namely, Intrex indexing.)

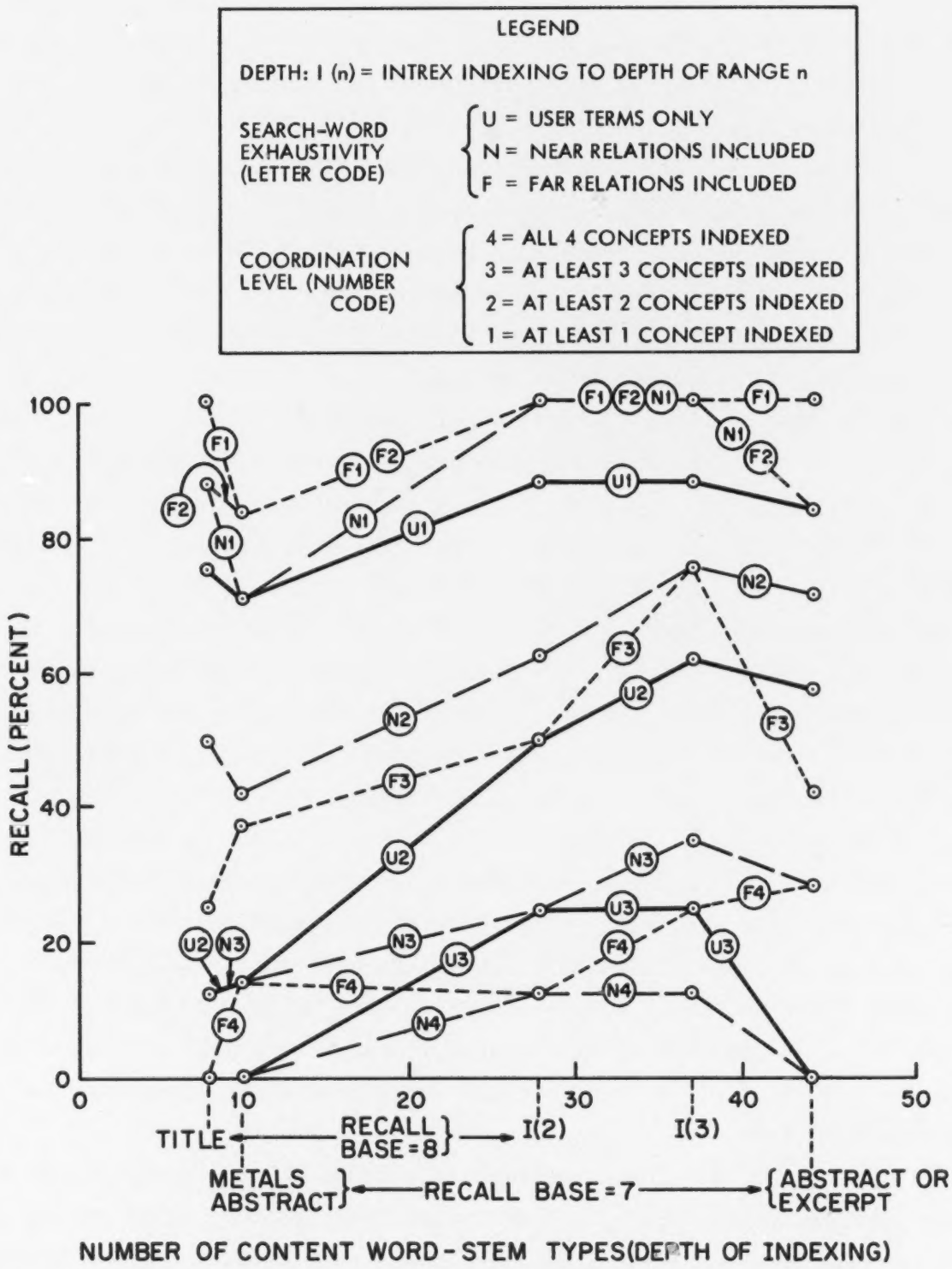


Fig. IIB-4 Recall as a Function of Indexing Depth and Type for Several Levels of Concept Coordination and Search Exhaustivity

Two other factors appear to be operative in the shape of the recall vs. depth curve. Firstly, a more or less obvious arithmetic fact, the higher the title-word recall, the less chance to avoid a diminishing returns (negative second derivative) effect. Secondly, there is what we might call the search exhaustivity factor. This has to do with the number of terms combined by Boolean ORs. The hypothesis is that the greater the exhaustivity factor the more the diminishing returns effect is likely to hold. The intuitive rationale for this hypothesis is fairly clear: the reason coordination seems to countermand the diminishing returns effect is the need to match on the greater variety of related words found in the deeper indexing: however, if the search strategy is such that most of the relevant words have already been included in the search statements, then going to deeper indexing will not help that much.

A quantitative measure of these factors may be obtained by considering the second derivative of the recall/depth curve as a measure of the degree to which the curve bends up or down as it is extended to the right (i.e., as depth increases). Table IIB-5 lists this measure for the different levels of coordination and relatedness.

Table IIB-5

Diminishing-Returns Factor (Second Derivative of Recall/Depth Curve) for Different Coordination and Exhaustivity Levels

Search Exhaustivity Level	Coordination Level			
	4	3	2	1
Single Terms (no disjunction)	-	-	-	- 0.71
All User Terms (U)	-0	+0.78	0	- 9.0
All User Terms and Near Relations (U,N)	+0.39	-0.78	- 5.5	-10.6
All Terms (U,N,F) (Full Exhaustivity)	+0.78	-1.56	-10.6	-12.5

The second derivative is approximated by taking the difference between the slope of the line connecting the recall values for title and for all Intrex terms and the slope of the line connecting the origin and the recall value for the title. The value for the coordination level of 1 and exhaustivity of single

terms is taken from Fig. IIB-4. It is reckoned that the average value curve given in this figure represents the cumulative normalized sum of individual searches on the separate terms.

In Table IIB-5 we see that, with two exceptions, all variation in terms of decreasing exhaustivity or increasing coordination level tends to increase the second derivative (i.e., lower the diminishing-returns effect). The two exceptions relate to U4, which is a degenerate case because recall is uniformly 0, and the anomalous relation between N4 and F4. Thus we have a rather striking overall agreement between the data for this experiment and the hypothesis given above.

A further hypothesis, suggested by this analysis, that can be postulated concerning the shape of the recall/depth curve, is that the diminishing-returns effect is enhanced as the level of relevance for a document to be considered relevant is increased. This seems to follow intuitively from the idea that if a document is considered more relevant, then it is more likely that the matching terms will be found in the more important range numbers and so there is relatively less need to go to deeper indexing. This hypothesis was not directly tested here although, as explained above, the diminishing-returns curve for single terms, as shown in Fig. IIB-3, seems to be a measure of how much this factor has effect.

Precision Analysis. The strategies formulated above were also subjected to a precision analysis. ES 12 was not available to give relevance judgements for this extended analysis so the Intrex analyst himself made relevance judgements on those documents which had not already been judged by ES 12 in previous experiments. The very extensive interviews with ES 12, together with the fact that analyst judgements proved compatible in terms of recall with the recall figures derived for the 8-document recall base, leads us to believe that the analyst judgements, on the average, are sufficiently reliable for this purpose.

The results of the precision estimation for various strategies and component strategies are shown in Table IIB-6. In some cases where the number of documents retrieved was large, not every document was checked for relevance; the precision figures in those cases result from an estimation based on sampling. Each strategy is defined in terms of which concepts are being coordinated, the depth of Intrex indexing being considered (by range number), and the level of search-term exhaustivity employed in the strategy.

In an analysis of the effect of coordination level it is appropriate to compare the strategies F4, F3, F2, and F1, which are all for full exhaustivity and depth of indexing and differ only in the number of concepts coordinated. Actually,

Table IIB-6

Precision Figures for Various Strategies for Problem of ES 12

Search Strategy					Number of Documents		
Label#	Coordination Level	Concepts	Depth of Indexing*	Search Exhaustivity+	Retrieved	Relevant	Precision (Percent)
F4	4	all	all	F	20	16	80
F3	3	all	all	F	225	62	27
dF3	3	all	all	F	205	46	21
ABF2	2	A,B	all	F	215	44	20
dABF2	2	A,B	all	F	111	6	5
CDF2	2	C,D	all	F	198	49	25
dCDF2	2	C,D	all	F	57	9	10
dF1	1	A	all	F	970	0	0
F4R2	4	all	range 2	F	9	9	100
N4	4	all	all	N	6	6	100
ABCF	3	A,B,C	all	F	40	29	73
ABCFR2	3	A,B,C	range 2	F	26	23	88
ABCFR1	3	A,B,C	range 1	F	7	5	71
ABCN	3	A,B,C	all	N	27	22	82
ABCU	3	A,B,C	all	U	11	10	91
BCDF	3	B,C,D	all	F	32	17	53
BCDN	3	B,C,D	all	N	11	8	72

The choice of labels was based on mnemonic devices to reflect the four strategy features as given in columns 2 through 5. Also, a lower case "d" at the start of a label indicates a "differential" strategy as explained in the text.

* Depth of Indexing: all index terms (full depth) or only those to the depth of the given range level. For example, the notation "range 2" alongside strategy F4R2 indicates that range 2 terms plus the more important range 1 and title terms are considered but not the less important range 3, range 4, and range 0 terms.

+ Search Exhaustivity: U = user terms only; see Table IIB-4
N = user terms and their near relations
F = user terms and their near and far relations

it is even clearer if we focus on the differential components of these strategies. Thus, for example, strategy dF3 includes only those documents retrieved by strategy F3 that were not retrieved by strategy F4 and dABF2 includes only documents indexed under both concepts A and B but not under concepts C or D. Note that on a coordination level of two, just two of the six concept-pair combinations were sampled and on level one just one of the four single concepts was sampled. From the figures we see that decreasing the coordination level from 4 to 3 increases the number of relevant documents retrieved by a factor of just under four at a cost of about 10 times the number of documents retrieved. Total precision drops from 80 to 27 percent with the incremental precision of the newly retrieved documents (dF3) being just 21 percent. Going to coordination-level two we see small increments in recall and we find precision plunging to values between 0.05 and 0.10 for the two cases considered. Other observations suggest that the recall level of F3 is about 75 percent and that of F2 is close to 100 percent. Since 893 documents are retrieved on the combined F2 strategy, this would give an overall precision of about nine percent for that strategy. A sampling of documents retrieved under the single concept A but under no other concept found no additional relevant documents.

In analyzing the effect of depth of indexing, we consider strategies F4, F4R2, ABCF, ABCFR2, and ABCFR1. Going from all index terms to just up to range 2, we increase the precision from 80 to 100 percent at a cost of a 55-percent drop in recall (from 20 to 9 documents). In going from complete indexing to range-2 indexing for the strategy that coordinates the 3 concepts A, B, and C at full exhaustivity, we increase the precision from 73 to 88 percent at a cost of a 20-percent drop in recall (29 to 23 documents). However, an anomalous result occurs when indexing is further restricted to range 1: precision drops to 71 percent while recall is sharply cut from 23 to 5 documents. Clearly, insisting on higher range-level matches is no guarantee of improved precision, while a drop in recall, sometimes drastic, is the inevitable concomitant.

In analyzing the effect of search exhaustivity, we consider strategies F4, N4, ABCF, ABCN, ABCU, BCDF and BCDN. Employing only user terms and their near relations at coordinate-level 4, instead of all search terms, raises the precision from 80 to 100 percent at a cost of a 70-percent drop in recall (from 20 to 6 documents). In the same transition process for the strategy that coordinates concepts A, B, and C, precision increases from 73 to 82 percent at a cost of a 24-percent reduction in recall (29 to 22 documents). If, for this same

strategy, we now employ only user terms, precision further increases to 91 percent but recall drops 55 percent (from 22 to 10 documents). A similar analysis for the strategy that coordinates concepts B,C, and D finds precision going from 53 to 72 percent and recall dropping 53 percent (from 17 to 8) when far relations are dropped from the search terms.

In summary, the coordination level is seen to be a very important factor for precision as well as for recall. Depth of indexing seems, in general, to have relatively small effect on precision compared to its often large effect on recall. Similarly, the level of search exhaustivity in this problem seems to have only a small or moderate effect on precision whereas its effect on recall is often quite large. Obviously, the effect of search exhaustivity on precision depends on how irrelevant the individual search terms become as the level of exhaustivity increases. In this case, it is felt that most of the terms brought in at higher levels of exhaustivity are still quite close to the user terms and hence the relatively small drop in precision is not unexpected.

A good search strategy for this problem would seem to involve a coordination level of three out of the four concepts and the greatest level of search exhaustivity, that is, the strategy labeled F3. This strategy gives high recall (about 75 percent) and moderate precision (about 27 percent). Recall would be considerably higher, probably close to 100 percent, if a higher relevance threshold were taken. The recall base, being based at least in part on the bibliography of ES 12, included some items of only background importance without high relevance to the central theme of the thesis, and it was mainly these types that were not retrieved by strategy F3.

Conclusions. A set of hypotheses has been propounded and partially verified for explaining the factors affecting the shape of the curve representing recall as a function of depth of indexing. In particular, the diminishing-returns effect (reduced second derivative of the curve) is lessened under the following conditions:

1. As level of coordination (ANDing) of search strategy is increased.

2. As level of exhaustivity (ORing) of search strategy is decreased.

3. As level of title-word recall is decreased.

4. As level of required relevance is decreased.

Specific effects in the problem analyzed here have also been noticed in other problems we have studied. In particular, coordination level has a very strong effect in both recall and precision. Depth of indexing has a relatively small effect on precision compared to its large effect on recall. Search exhaustivity may have great or small effects on the various retrieval-effectiveness parameters, depending on the vocabulary requirements of the problem at hand. General word linking within terms — that is, the use of the Intrex command WITH instead of AND — adds little to precision but seriously deflates recall; however, in particular situations (for example, "high-temperature") even the stronger, word-adjacency requirement is needed for adequate precision.

The quantification of these factors bears important implications, of course, for the choice of the appropriate indexing depth in retrieval systems. In particular, we have identified a class of problem, associated with high levels of coordination, in which a combination of deep indexing and high search-term exhaustivity is required to achieve satisfactory levels of retrieval effectiveness.

Results of this analysis indicate that abstract-word indexing will probably be somewhat less effective than full Intrex indexing, especially on a per-word basis, and perhaps only as good as Intrex range-2 indexing. Two factors adversely affecting the performance of abstract words are:

1. Some abstracts are only fair to poor indicators of subject content. Excerpts used as abstracts are especially suspect.

2. Abstracts contain a higher proportion of non-content words than Intrex indexing.

The use of the number of content word-stem types as a measure of depth of indexing has been better established in this analysis.

It appears that Metals Abstract indexing — and other controlled-vocabulary indexing — will at times be somewhat less good than title-word indexing due to, as in this case, failure to include coverage of certain important specific concepts in the topic. This is indicated for certain topics despite the fact that MA indexing on the whole may be likened to title plus keyword or about at the level of Intrex range 1.

This analysis gives further supporting evidence to the hypothesis that controlled vocabularies do not in practice exhibit very much of the supposed good recall effects of normalizing vocabularies. This appears to be largely attributable to the fact that only a small number of terms or "pigeon-holes" can, in

practice, be assigned to each document and so, while a given term may be applicable, it simply will not be applied. Some minor utility in the morphological control over word forms in MA (for example, "aluminum" = "aluminium") is found in this study, but semantic control (for example, "high temperature" = "elevated temperature") appears to have had very little success.

CATALOG-INDICATIVITY EXPERIMENT

The catalog-indicativity experiment is designed to test the effectiveness of different types of catalog information as indicators of the value of documents to users. To date fifteen experimental subjects (ESs) have participated in the Series B part of the experiment; eleven have completed the experiment, and four are currently in progress. All fifteen ESs are at M.I.T. Two are professors, one is a post-doctoral assistant, eleven are graduate students working on their doctoral theses, and one is an undergraduate. The catalog-indicativity experiment will be completed as soon as four more undergraduates and two more professors have served as experimental subjects — thus bringing the total number of undergraduates to five and the total number of professional researchers also to five, a minimal number for valid comparisons among groups. For a complete description of the methodology of the experiment, see the 15 March 1971 Semiannual Report.

Since the last Semiannual Report we have, in addition to continuing to run the experiment, made an analysis of three major areas being investigated by the experiment: indicativity of catalog fields, reasons offered by ESs for evaluating documents as other than 'highly useful', and post-experimental interview responses of ESs. Discussions of the findings in these three areas appear below.

Indicativity of the Catalog Fields. Results concerning the indicativity of the four main content-indicating fields (title, abstract, all subject-index terms and matching subject-index terms) have not changed significantly since the last Semiannual Report. Our data have been increased by two ESs who have recently completed the experiment, but the initial trends reported last time still continue: abstracts and subject-index terms are the most indicative of the four fields, matching terms are somewhat less indicative than abstracts and index terms, and titles are least indicative.

In Series B we have gone beyond the basic evaluation of the indicativity of the four main content-indicating fields to attempt to evaluate all of the

catalog fields and to better understand the ES's reasons for making his choices. A preliminary accounting of this extended analysis for the first 11 ESs is given below. Completion of the basic and extended analyses will follow the conclusion of the running of the 20 ESs.

In order to ascertain the impression ESs had about the utility of the catalog fields — as opposed to the indicativity scores resulting from concurrence of catalog field ratings with text ratings — we asked ESs their opinions directly. First, each ES was given a list of the 54 catalog fields and a general description of the kinds of information which typically appear in these fields. The ES was asked to place check marks beside fields which contain the kind of information that would have been helpful in making document evaluations — double check marks were suggested for those fields that would have been especially helpful. Second, each ES was given the catalog records of three documents and was requested to place single or double check marks beside fields which contain information that would have been helpful in evaluating those particular documents.

The most helpful fields are listed in order of helpfulness, on the basis of general descriptions and appearance in actual catalog records, in Tables IIB-7 and 8, respectively. Fields which were in the list, but were never checked, do not appear in either of these tables. A field which has the same total percent of checks as another field, but a greater percent of double checks, is considered to be more helpful; such fields are marked with asterisks. In Table IIB-7, the percentages of fields checked on the basis of general descriptions are calculated on a base of eleven possible checks. In Table IIB-8, the percentages of fields checked on the basis of actual catalog records are calculated on a base of the number of occurrences of each field in the catalog records; since each of eleven ESs was presented three catalog records, the maximum number of occurrences of any field is thirty-three. The actual number of times each field occurred in the catalog records is recorded in Table IIB-7.

It is to be noted that even with such a small number of ESs with similar research topics, and with a data base primarily centered on one document type, a large number of fields — 63 percent by general description and 56 percent by actual fields presented — were checked as helpful; and fields which were not checked were frequently fields which we did not anticipate would be helpful to users — for example, cataloger (field 2). We would expect a larger sample of ESs, with more diverse research topics and in the presence of a more varied data base, to find an even larger percentage of fields helpful.

Table IIB-7

Percent of Fields Checked by ESs,on Basis of General Descriptions, as Helpful in Making Document Evaluations

Catalog Fields (Field Number)	Total Percent Checked
Abstract (71)*, Subject-index Terms	100
Text (90)*, Title (24), Excerpts (70)	91
Author (21)	82
Purpose (65)*, Match (74)	73
Language (36)	56
Affiliation (22), Language of Abstract (37)	45
Publication Date of Book or Report (29), Format (31), Table of Contents (67), Features (68), Bibliography (69)	36
Citations (80)*, Approach (66), Reviews (72)	27
Library (11), Pagination (32), Thesis (43), Journal Location (47), Normal (76), Comments (85)	18
Illustrations (33)*, Supplement (41)*, Holdings (12), Main Entry (20), Coden (25), Publisher (27), Medium (30), Series (38)	9

Table IIB-8

Percent of Fields Checked by ESs,on Basis of Occurrence in Catalog Records, as Helpful in Making Document Evaluations

Catalog Fields (Field Number)	Number of Occurrences in Catalog Records	Total Percent Checked
Title (24)	7	100
Abstract (71)	29	97
Subjects (73)	33	88
Author (21)	6	83
Contents (67)	12	75
Excerpts (70)	3	67
Purpose (65)	33	55
Features (68)	9	44
Language (36)	33	33
Language of Abstract (37)	24	33
Affiliation (22)	32	31
Illustrations (33)	31	10
Approach (66)	33	9
For Whom Chosen (2)	33	3

From Tables IIB-7 and 8 it appears that the fields (other than the main content-indicating fields) which are most helpful are fields 21 (author), 36 (language), 65 (purpose), and 67 (table of contents). Each of these fields was checked more than fifty percent of the time on the basis of general descriptions or catalog records or both. Other fields are very helpful to some ESs but not helpful in the majority of cases. In general, ESs tended to check fields less often in the catalog records (e.g., field 36 was checked 55 percent of the time on the basis of general descriptions but only 33 percent of the time in the catalog records). The decrease is, however, understandable; a field was checked in the general descriptions if at all helpful (one check per one occurrence), but perhaps checked only once out of three occurrences in the catalog records if at all helpful. There were exceptions to this general rule, though; field 67 (table of contents) was checked 36 percent of the time in the general descriptions and 75 percent of the time in the catalog records. It is clear that the ESs could not tell from the general descriptions alone how helpful the contents field would be; in the future we must focus on the question of how to encourage users to use such fields — fields which are actually very helpful but which users are not likely to pick on their own.

Many fields (11, 12, 25, 27, 29, 32, 38, 41, 72, 76, 80 and 85)* were checked in the general descriptions but were not checked in the catalog records because, being primarily for document types not in the data base, they never appeared there. (Even though there are fifty-four catalog fields, the average number of catalog fields in each catalog record for the eleven ESs was only 23.3). We have a rough idea of the usefulness of these fields from the number of times they were checked in the general descriptions; but we need to present to ESs some actual information from these fields in order to get a more accurate measurement of the real usefulness of these fields.

Fields 20 (main entry), 30 (medium), 31 (format), and 69 (bibliography) were checked in the general descriptions but were never checked in the catalog records, even though they did occur there. It appears that these fields, even though they may sound helpful to the ESs on the basis of the general descriptions, are in actuality not very helpful. Fields 43 (thesis) and 47 (location) were also checked in the general descriptions but not in the catalog records; however,

---

* See Table IIB-7 for titles of these fields.

no conclusions can yet be drawn from this since field 43 occurred only twice in the catalog records,and field 47 occurred only six times (in the catalog records for two ESs).

Reasons for Evaluating Documents as other than 'highly useful'. In the catalog-indicativity experiment we were interested in exploring the reasons why ESs judge documents to be other than 'highly useful' to them in their research. Thus we asked each ES to cite certain reasons to explain why he gave an evaluation of '2' ('The article is somewhat useful') or '3' ('The article is not useful') to any document. Table IIB-9 indicates the percent of cases in which certain reasons were offered to explain '2' and '3' ratings. The percents total more than 100 percent because occasionally two or three reasons were offered to explain single ratings. The data for Parts I and II are drawn from 13 and 12 ESs, respectively; and the percents are calculated on the basis of 842 and 239 total reasons cited, respectively.

The only reason which was never used to explain ratings of '2' and '3' was: 'The article is experimental, rather than theoretical'. Undoubtedly that reason was never used because all the thesis students are working on experimental theses; and the other three ESs were interested primarily in experimental articles. The largest percent of reasons had to do with the topical irrelevance of the articles (the first three reasons listed in Table IIB-9); 86.5 percent of the documents in Part I and 91.8 percent of the documents in Part II which were evaluated as '2' or '3' were evaluated so because of their irrelevance to the ESs' topics. The relatively small incidence of reasons cited from the "Nature of Article" category compared to the "Topical Relevance" category might suggest a rather small utility of most catalog fields other than the main content-indicating fields. However, here again, we must urge caution in drawing conclusions because of the rather homogeneous nature of the data base, namely, mostly recent, high-quality, professional journal articles.

Reasons in the 'Other' category were cited to explain seven percent of the '2'- or '3'-rated documents in Part I and three percent of the '2'- or '3'-rated documents in Part II. An analysis of the comments in this category indicates that only four of the seven percent in Part I and one of the three percent in Part II truly fit into this category; the remaining comments in the 'Other' category fall more appropriately into categories (5),(8), and (11) in Table IIB-9. The reason most frequently cited in the 'Other' category was that the information was insufficient.

Table IIB-9

Breakdown of Reasons Explaining Documents as Other Than 'Highly Useful'

Reason	Part I (Evaluation of Catalog Field Information)	Part II (Evaluation of Full Texts of Documents)
Insufficient Relevance of Article		
(1) The article is not at all relevant to my research topic.	42 percent	39 percent
(2) The article is only indirectly relevant to my research topic.	33 percent	40 percent
(3) Only a small portion of the article is relevant to my research topic.	14 percent	16 percent
Nature of Article		
(4) The quality of the article is inferior.	0 percent	1 percent
(5) The article's treatment of the topic is too superficial or elementary.	under 1 percent	2 percent
(6) The article is not fully understandable, or it lacks textual clarity.	under 1 percent	0 percent
(7) The article contains no new information, does not move beyond the current state of knowledge in the field.	under 1 percent	under 1 percent
(8) The results of the article are outdated.	1 percent	3 percent
(9) The article is experimental, rather than theoretical.	0 percent	0 percent
(10) The article is theoretical, rather than experimental.	4 percent	5 percent
(11) The orientation of the article is not appropriate (it is overly mathematical, is about the wrong material, etc.) — please specify.	6 percent	13 percent
Other		
(12) Please comment.	7 percent	3 percent

Post-Experimental Interview. After they complete Parts I and II of the experiment, ESs are interviewed to obtain information for use in interpreting experimental results.

From such interviews we find, for example, that users seem about evenly divided between those who look at half of the article or more (5 ESs), and those who look at less than half (6 ESs). Most (10 out of 11) say that they only scan those parts of the article they look at. More users (10 out of 11) base their evaluation, at least in part, on the conclusion or summary than on any other part of the article. The introduction was used for evaluation by more than half of the ESs (7 out of 11). Only 4 out of 11 said that they paid special attention to the illustrations and none said they used the bibliographic references for this purpose.

These results are generally consistent with those found in the text-access experiments described in a preceding section and help to develop the pattern of text usage which will help us analyze the optimum configuration for a text-access system. Many other questions were part of the interviews; other results from the interviews will be presented when additional analysis on the Series-B experiments is completed.

C. ECONOMIC ANALYSIS

Staff Members

Mr. M. K. Molnar
Dr. C. W. Therrien

Graduate Student

Mr. H. V. Jesse

SUMMARY

Information-retrieval-system modeling has continued. Results of simulated operation of an I-R system for markets with varying degrees of elasticity indicate that most stable operation is realized when the market is inelastic.

In order to estimate the costs of hardware needed to support I-R system operation, topics related to the design of an operational system were considered. Suggested hardware configuration and software organization are presented in this section.

Topics related to design of an information-retrieval network are under investigation. It is desired that users be able to access a number of different remote data bases and to conduct online searches on these data bases in a single command language. Topics related to communication requirements, language requirements, and cost optimization are being considered.

MODELING OF INFORMATION-RETRIEVAL SYSTEMS

Research has continued on the economic modeling of information-retrieval systems. Attention was focused on the operational characteristics of an I-R system through study and simulations centered on the dynamic model cited in Fig. IIC-4 of our 15 September 1971 semiannual report. Computer programs for this model have been refined and the revenue function, derived from stochastic analysis of the service process, was integrated with the rest of the model. The revenue function, although not expressible in closed form, is nonetheless quick to compute on a digital computer, and so warrants incorporation into the model.

An important feature of the dynamic I-R system model is the user demand function. This function, like all demand functions encountered in economics, relates price (of service) to quantity (of service) that users are willing to purchase at the given price. As such, appropriate variables for the demand curve are w, the average hourly charge to a user, and f, the average number of I-R sessions required by a user per month. The user demand curve, when plotted with the axes as shown in Fig. IIC-1, slopes downward and to the right and intercepts the f-axis at a point f_o which represents the number of sessions users require when there

is no charge for service. The demand curve of Fig. IIC-1 is different from that generally encountered in economics in that, because of the saturating nature of the service process, the number of sessions demanded is not necessarily the number of sessions users actually receive. Thus total revenue is a nonlinear function of the number of requested sessions. The function that expresses this

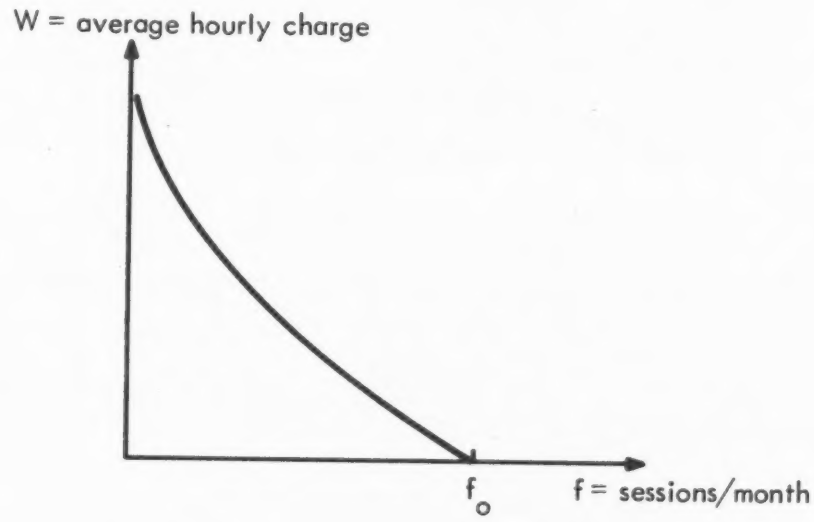


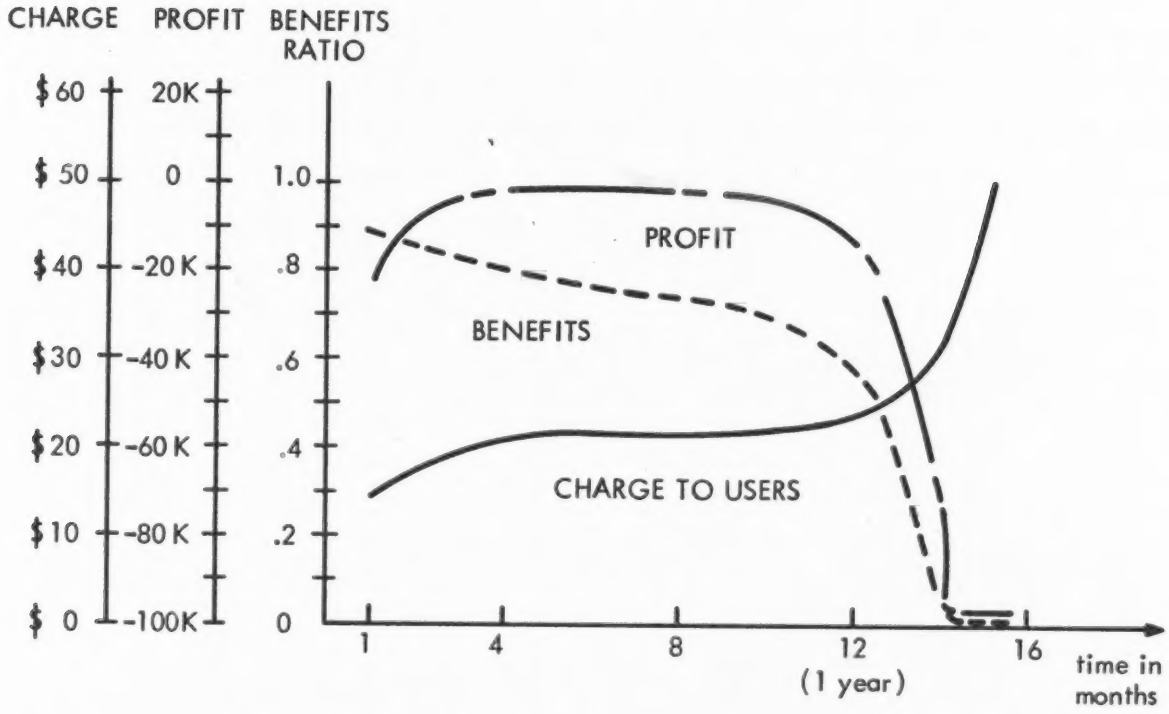
Fig. IIC-1 A Demand Curve for I-R Service

nonlinear relation can be applied to the user demand function of Fig. IIC-1 to derive another demand curve which relates the number of hours of service supplied to users, to the hourly charges (w). An important property of this new demand function is its elasticity, which measures change in total revenue for a small change in price. Qualitatively, a demand is said to be 'elastic' when a small decrease in price results in an increase in total revenue. A demand is said to be 'inelastic' when a small decrease in price decreases total revenue.* The elasticity of the demand is an important factor in the dynamic performance of any user-supported I-R system.

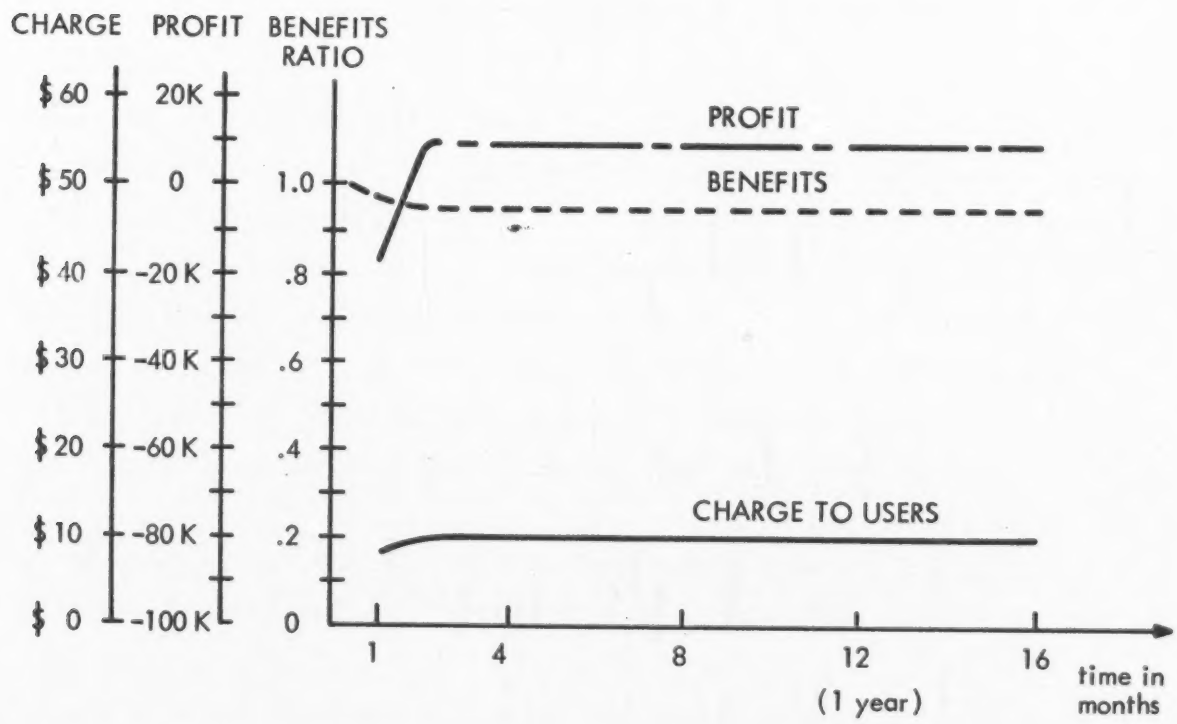
Simulations of I-R system performance were carried out for demand functions with varying degrees and regions of elasticity. It was assumed that the

---

*If p and q are the price and quantity variables, then the elasticity is defined by $E = \frac{dq}{dp} \Big/ \frac{q}{p}$. The ranges of E corresponding to an elastic and an inelastic demand are respectively $(-\infty, -1)$ and $(-1, 0]$.



(a) Users require maximum of two sessions per month. Number of simultaneous users is 30.



(b) Users require maximum of four sessions per month. Number of simultaneous users is 60.

Fig. IIC-2 Simulation of I-R System Operation

economic objective was to operate the system on a no-loss basis. Basically the results are these:

- Any policy that seeks to overcome financial loss by adjusting user charges over a portion of the demand curve where demand is elastic results in economic instability and ultimate catastrophe.
- An I-R system can and should overcome net loss by adjusting charges whenever the user demand is inelastic.

Thus, by observing incremental changes in revenue attributable to small price changes, a system manager can properly adjust charges for best operation of the system with respect to his clientele of users. Examples of simulations illustrating the above two points are shown in Fig. IIC-2. The dotted line labeled "benefits" is a combined measure of demand and system effectiveness in filling the demand. Specifically this measure is defined by

$$\text{benefits ratio} = \frac{f}{f_o} \cdot \text{service index} = \frac{f}{f_o} \cdot \frac{R}{R'}$$

where f and f_o are explained above, R is the total revenue and R' is the maximum possible revenue. Note in Fig. IIC-2(a) that in the range of zero to $20 per hour the demand curve is inelastic and charges can be increased (over a period of eight to ten months) to minimize the loss. However, further attempts to increase charges bring charges into a region where demand is elastic and result in rapid diversion. Figure IIC-2(b) depicts a situation where the demand is inelastic in the region of operation and attempts to increase charges lead immediately to satisfactory profitable operation. This system behavior can be quantified mathematically.

On the basis of the foregoing studies it is clear that for stable no-loss operation of an I-R system it is preferable to have a market whose demand curve has a relatively large inelastic region. This has implications for the manner in which I-R services should be supported. If users pay for I-R service as individuals, then one would expect the demand curve to be largely elastic, i.e., relatively sensitive to price. However if charges for I-R service are underwritten by organizations, especially if such charges become a "line item" in research budgets, for example, the demand might be expected to be relatively insensitive to price changes or inelastic. It is thus reasonable to conclude that even if I-R service charges can be brought within the reaches of the

individual user, the institutional market is probably a market that is more stable and so better matched to this type of service.

DEDICATED INFORMATION-RETRIEVAL SYSTEM DESIGN

In order to estimate costs involved in providing computer facilities to support an operational information-retrieval system it was necessary to devote some attention to topics related to the actual design of an operational I-R system. Although some consideration of topics related to operational-system design has been going on concurrently with economic analysis for over a year, additional attention has been given the topics in the last few months.

An operational I-R system of the type considered here is centered on a dedicated computer. We have made the assumption that this computer should be capable of hosting at least 30 to 60 simultaneous on-line users and provide access to a data base of approximately one million entries. In addition, the system should be able to process requests by users on a deferred basis, and process SDI profiles during updating of the data base. A block diagram for the hardware of a system with these capabilities is shown in Fig. IIC-3. On-line users

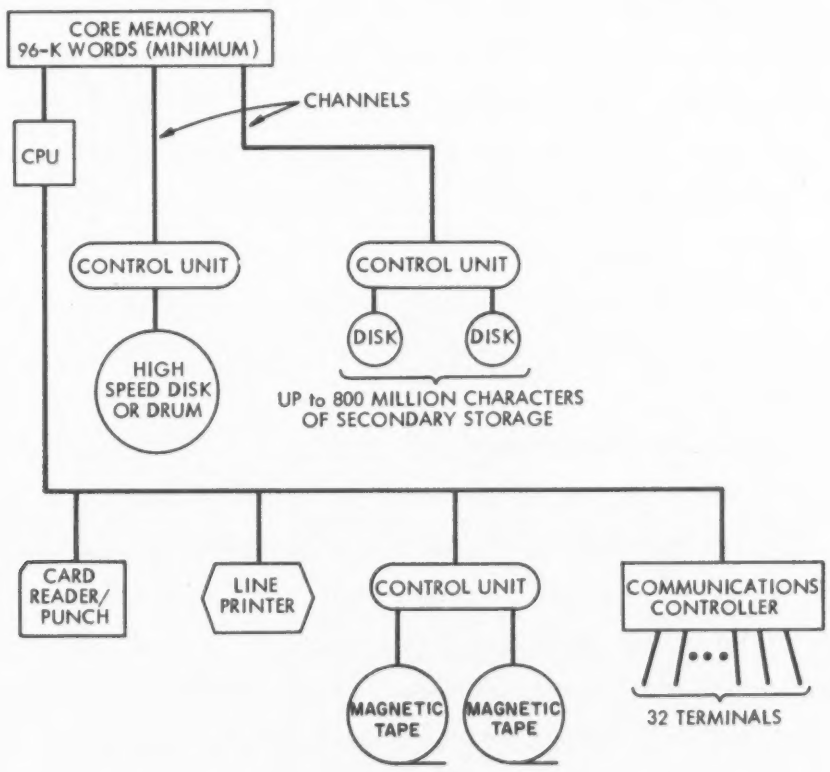


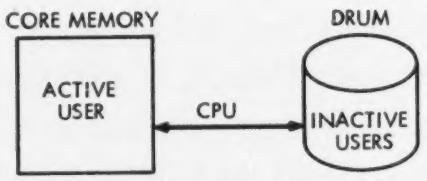
Fig. IIC-3 An Interactive Computer System Dedicated to Bibliographic Information Retrieval

communicate with the central-processing unit (CPU) via the communication controller to which phone lines are connected. A high-speed disk or drum is provided for swapping users' data in and out of core in the process of time-sharing. The bulk of the data base (catalog records and inverted file) is stored on the high-density disk units. Assuming a catalog record size of 400 to 600 characters and allowing a 30 to 50 percent overhead for inverted files and directories, we find that the disks should have capacity of 600 million to 800 million characters of storage. Other peripheral devices including card reader/punch, printer, and tapes needed for operation of a machine this size are also part of the configuration. The size of core memory required is based on the proposed structure of the programs and stems from our experience with Intrex on the M.I.T. compatible time-sharing system (CTSS). On CTSS, time-sharing is accomplished by bringing a single user program into core memory, executing it for a short period of time, swapping the unfinished program to a drum, bringing in the next user's program, executing it, swapping it, and so on through the set of on-line users. The process is depicted in Fig. IIC-4(a). More modern computers have built into them a capability for maintaining more than one user program in core at a time. Thus, when one program execution is interrupted for some reason, the central-processing unit begins execution of another in-core program immediately. In the meantime, the interrupted program can be swapped to a drum or disk by means of a separate processor (usually called a channel). Thus, central-processor time is not wasted because of swapping (see Fig. IIC-4(b)).

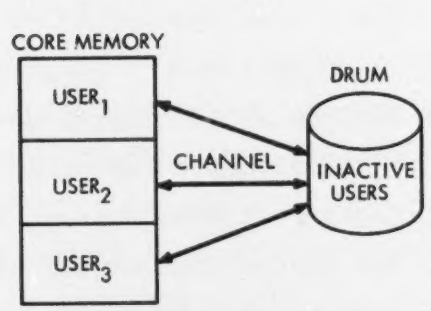
The process just described, called multiprogramming, is an order-of-magnitude improvement over pure swapping, as performed on CTSS. In the case of a dedicated information system, another order-of-magnitude improvement can be made via the concept of re-entrant programming. Note that when the configurations of Fig. IIC-4(a) and (b) are used for a dedicated system, all users are, in fact, running the same program (the retrieval program). Thus it is wasteful both with respect to core memory and with respect to swapping effort to maintain multiple copies of the program, one for each on-line user. Re-entrant programming permits organization of the program into a "pure" code segment shared by all users, and impure "data" segments corresponding to each user, as shown in Fig. IIC-4(c). This organization requires less core storage, less swapping, and thus yields increased efficiency.

The description of software organization has so far been presented as if the retrieval programs were under control of an existing supervisor or monitor

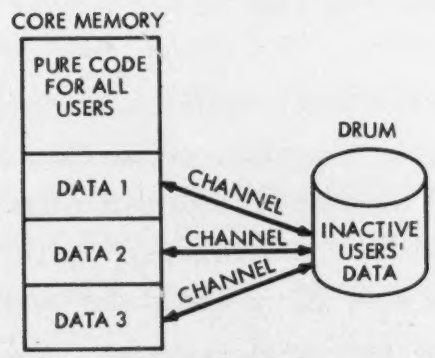
program (not shown in Fig. IIC-4) that handles terminal communication, schedules execution of user requests, initiates swapping, and so on. Since, in the dedicated mode of operation, both the supervisor and the pure-code portion of the retrieval programs are core resident, these two programs can be replaced (at least conceptually) by a single system program that performs both supervisory and retrieval



(a) Swapping System



(b) Multiprogramming System



(c) Multiprogramming System with Re-entrant Code

Fig. IIC-4 Levels of System Structure for Computer Dedicated to Bibliographic Retrieval

functions. Practically, however, there are reasons to prefer modular separation. First, most functions performed by the supervisor are fundamental to the operation of any on-line time-shared system. The art of supervisor design and implementation has developed to a point where information-system programmers should best build

upon existing designs. There is no need to "reinvent the wheel" (although we might add some springs to the suspension system). Second, and probably more importantly, the existing monitors provide an environment where general-purpose time-sharing operations can coexist with dedicated-system operation. Thus, needed software repairs and modifications can be made on-line while the system is supporting information-retrieval operations. In this case the retrieval programs are core-resident at all times except during the relatively few time slices allocated to system programmers making repairs. Thus retrieval-system users should experience only very slight degradation in performance.

An item as important as the operating system is the file system. Just as the operating system should permit programs to have the desired structure in core storage, so should the file system permit data to have a known desired structure in disk storage. Our conclusions regarding the file-system structure for a dedicated I-R computer remain generally in agreement with those of R. L. Kusik who designed a file structure for information retrieval on the IBM 360 Model 67.* To state these briefly, it is important that the file system permit blocks of data (records) to be placed in physically adjacent positions on the disk. This structure eliminates the overhead of multiple "seeks" or mechanical efforts to locate the records comprising a large file. In addition, it should be possible to bypass most parts of the file system and address data directly by a location on the disk. This ability facilitates construction of the distributed directories to data, as proposed by Kusik. Once so constructed, files can be read in a near optimal manner through the standard facilities of the file system.

Typical medium-scale computers that provide capability needed for a dedicated I-R system of the assumed capacity can be purchased for about $1 million to $1.5 million, depending on the manufacturer. Monthly lease costs are in the range of $26,000 to $33,000.

---

*R. L. Kusik, "A File Organization for the Intrex Information-Retrieval System", Technical Memorandum ESL-TM-415.

INFORMATION-RETRIEVAL NETWORKS

Introduction. Since study of information-system networking is a new topic for this project, it is appropriate to make a few preliminary comments about goals.

Several automated bibliographic information storage and retrieval systems have been developed in recent years. These systems have been motivated by the rapidly increasing abundance of literature, particularly of a scientific and technical nature, and have been made possible by the concurrent development of the computer with increased processing capabilities and high storage densities.

Although the development of I-R systems have been a major advance in the automation of information retrieval, problems with the technology remain which, if not solved, will tend to limit the usefulness of automated I-R systems. Among these problems are the following:

1. Data-base capacities are limited. A data base containing bibliographic information for 10^6 documents represents a large collection for effective on-line operation, yet it represents very few documents with respect to the total amount of published literature.*

2. The above data-base constraints necessitate that, in order to be most effective, a particular center concentrate upon one or a few related topic areas which can be expected at best to satisfy only a majority of the local user population.

3. A number of I-R systems have been developed which in their present form inhibit the interchange of information between centers because of differing file structures, differing search techniques, and physical separation.

4. The access of needed information by an individual user is impeded by political, procedural, and accounting considerations and the relatively high cost of data communications over voice-grade telephone lines.

A solution to these problems may ultimately lie in the construction of a large-scale information-retrieval network where users can access many data bases from a single host computer (and thus preferably interact in a single

---

*Retrieval systems that operate in a batch-processing mode can, of course, handle much more than 10^6 bibliographic references. These systems are inherently slow from the point of view of "turn-around" time and consume a large amount of computer-processing time.

language). The computer should in turn direct inquiries to the various data bases over high-speed lines and serve as a data concentrator.

The Network Concept. At least two distinct modes of operation can be distinguished in an online information retrieval system. One is the online interactive mode such as on the Intrex Retrieval System where the user engages in a dialog with the computer to search for documents of interest to him. A second mode of operation is the so-called "deferred search" mode where a request for information is submitted to the system from an online console for later processing. In this mode no interaction takes place except in setting up the request. The inclusion of this mode of operation serves to provide a less expensive service for those who have identified the area that is desired to be searched, and to smooth the stochastic characteristics of the computational load. Both types of operation can be incorporated into an information-retrieval network although for networking deferred searching is a task somewhat simpler than online interaction. Nevertheless online interaction is such a highly desirable mode of searching that it is indeed appropriate to consider the feasibility of interactive searching in the context of a network and to determine how it best can be implemented.

Attention has been focused on hardware configurations that might be used to support interactive retrieval through a network. For example, a very simplistic network might be as indicated in Fig. IIC-5 where many retrieval computers access a data base located in a remote storage device. Without networking, the controller for a storage device would interface directly with the channel of the computer system. The channel in turn has access to both the central processing unit (CPU) and core memory. Users are shown interacting with the CPU although they may in fact have access to the system through a channel or input/output bus. Special communication information processors (CIP) on either end of the communications link serve to provide buffer storage for queues, parallel-to-serial bit conversion, and other similar functions. This simplistic scheme has several disadvantages including the requirement of identical file and search systems at each center and high communication traffic, since all logical operations on lists are performed after transmission.

Because of these limitations another configuration has emerged as more promising, namely, the interconnection of locally complete I-R centers, as shown in Fig. IIC-6. In this configuration logical operations can be performed at the remote site as well as locally, thereby reducing communications traffic. Different centers can continue to specialize in particular subject areas, but users are

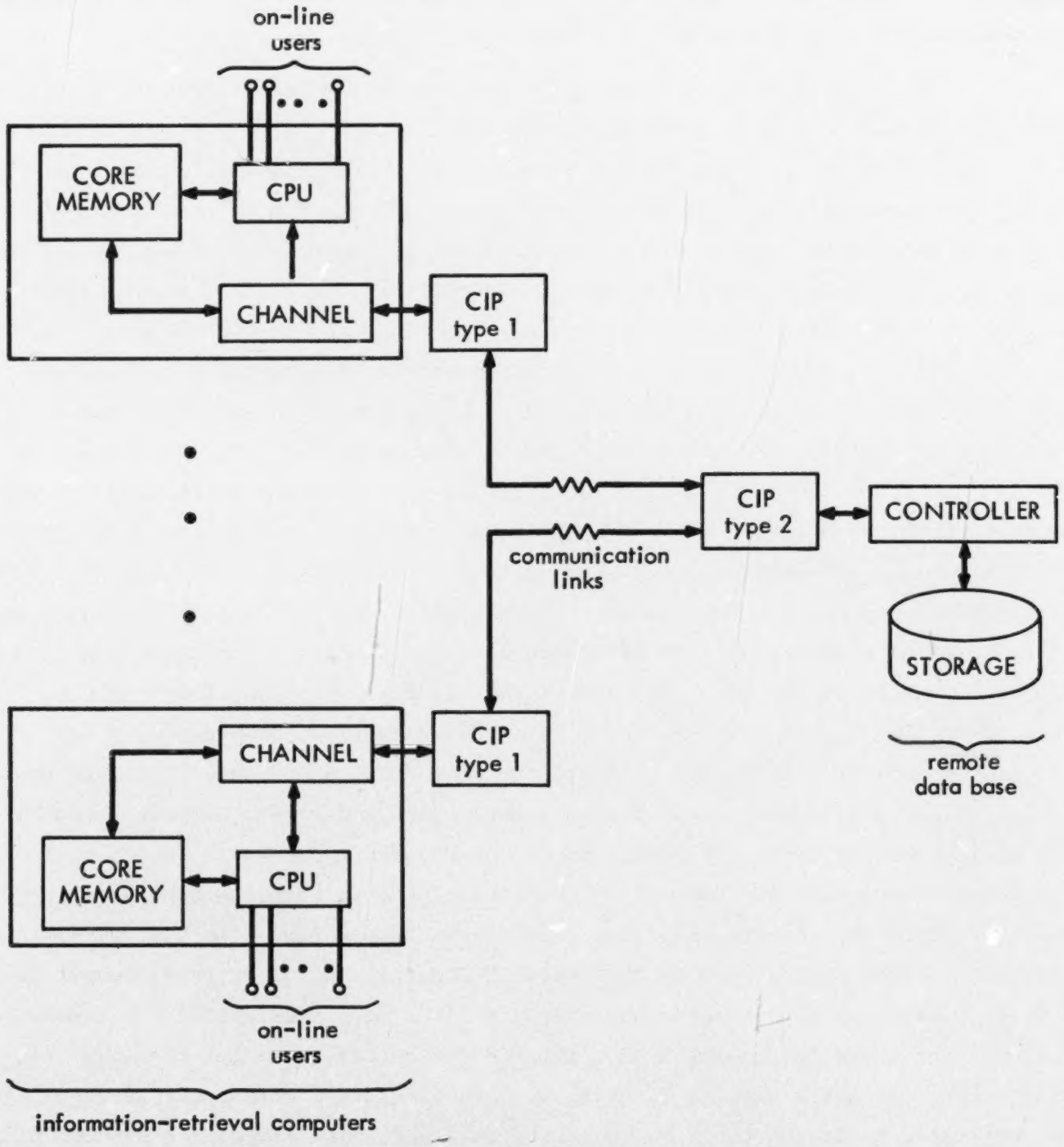


Fig. IIC-5 Information-Retrieval Computers Accessing a Remote Data Base

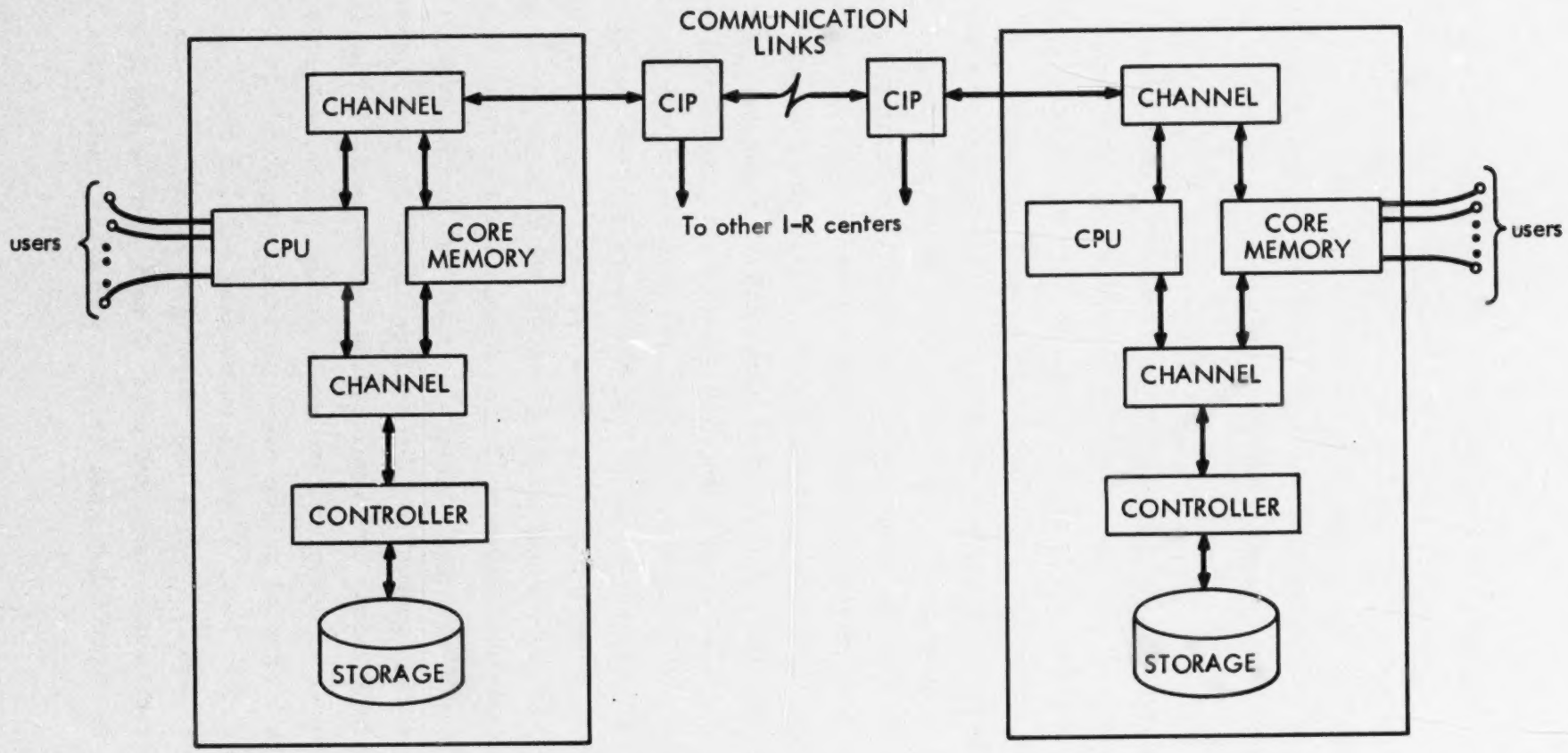


Fig. IIC-6 Networking Involving Locally Complete Information-Retrieval Systems

not constrained to their local data base. I-R centers need not change their file or searching structure but rather need merely provide (in the CIP) for translation of commands and search results into an appropriate language for communication.

Related to the implementation of a network of time-shared, locally complete I-R centers with specialized data bases are numerous technical aspects such as communications requirements and command commonality. These two topics are discussed below.

Communications Considerations. Important communications issues for interactive searching through a network are capacity requirements for a specified number of simultaneous users and a specified rate of requests, and expected message delays for given line capacities and network topology.

In order to determine feasibility of large-scale networking, attention has first been focused on the expected communication traffic. A model has been developed for the case in which two geographically separated interactive time-sharing systems are interconnected. Representative figures have been determined and channel capacities have been estimated. These results can be generalized to network topologies where there are no intermediate nodes and some symmetry. On the assumption of fully operational I-R centers with references to 10^6 documents in the data base, a user community of 12,000 persons at each center, and a symmetric traffic pattern in which half of the users at each center use a data base at a remote site, it was found that for a typical file and search structure the expected communication-channel traffic was 45 kilobits/sec in each direction. This result was computed assuming a rich dialog in which each user at every terminal issued two commands per minute and received, on the average, citations for 25 documents for each request to the catalog file. It was further assumed that the number of simultaneous users was 120, a number predicted by Goldschmidt* for suitably structured information-retrieval systems but a number about two times greater than the number of online users generally supported by ordinary time-sharing systems. Moreover, it was assumed that all 120 simultaneous users were using remote I-R centers. Hence the computed traffic of 45 kilobits/sec might represent an instantaneous peak and half of this traffic might be a typical average value. It is significant that the bit rates fall below 50 kilobits/sec since

---

*Goldschmidt, R. E., "File Design for Computer-Resident Library Catalogues", Electronic Systems Laboratory, M.I.T., p. 213. ESL Report ESL-R-451 (also PhD thesis)

50 kilobit lines are presently available from the common carriers on both a leased line and a dial-up basis. Thus it appears that information retrieval networking operations can effectively be supported by communication networks now in existence. It also shows that, at least with regard to the communication traffic, information-retrieval operations would be possible on the ARPA network whose computers are interconnected with lines of 50 kilobits/sec or greater capacity.

Currently the expected delays for message bits sent along communications lines with the above-stated loading are being investigated using a model developed by Kleinrock.* It should further be of interest to examine the stochastic behavior of both the delays and the traffic. If it can be shown that for suitable realistic conditions requests are short and can be considered to be statistically independent, a Poisson-like model may be appropriate. Otherwise, some other mathematical description of the process may be more aptly applied.

Searches Between Dissimilar Systems. In order to effect searches between information-retrieval systems with different command languages and different data base organizations, one could envision the incorporation of special modules in the network which translate a request issued in a user's local command language into a request that can be recognized by the retrieval system that is to process the search. Similarly, results of searching are translated by the module from a response in the search language to a response in the user's language. Although the scheme just described might well be feasible when only two information-retrieval systems are involved, the number of translation modules becomes too large to be practical when greater numbers of systems are considered. For n dissimilar retrieval systems, the number of distinct translation modules required is proportional to n^2. A much better alternative is to develop a single common I-R language into which each center would map its own language. A preliminary investigation of three interactive retrieval systems** shows that there is considerable overlap of functions and that development of a common language should indeed be possible.

---

*Kleinrock, L., "Analytical and Simulation Methods in Computer Network Design", Proceedings of the Spring Joint Computer Conference, 1970, AFIPS, Vol. 36, pp. 569-579.

**The three systems are Intrex, NASA's RECON, and SDC's ORBIT.

The most fundamental problem in bringing about communications between two dissimilar systems seems to be the mapping of a free vocabulary into a controlled vocabulary, or the mapping of one controlled vocabulary into another. The problem is not new, however, since any controlled-vocabulary system has been faced with the task of developing methods to help the user formulate his request in terms appropriate for searching. A technique currently in use is to display a portion of the thesaurus that most closely matches an initial free-vocabulary subject term typed by the user. The user may then select the appropriate search terms needed to retrieve documents in his area.

While preliminary investigation of the problem related to conducting searches between dissimilar systems indicates that such intersystem searches could be made possible, several detailed issues need further study. In particular, the structure of a common language suitable for intersystem communication must be defined together with suitable transformations of currently existing languages into the common language. A suitable method to aid users of the network in dealing with unfamiliar controlled vocabularies must be carefully worked out. Finally, a design of the modules that translate each retrieval-system language into the common language must be developed with estimates of both time-of-operation and core-memory requirements.

NETWORK COST OPTIMIZATION

In the design of an information retrieval network, communications costs are critical in determining optimal placement of I-R centers. The groundwork has been laid for an approach to network optimization.

Given a number N of information-retrieval centers geographically located in some way and a geographical distribution of users that are serviced by each center to a given level of satisfaction (say the service index is not less than 0.9) then one can compute a (minimum) total cost for supplying I-R service to users. Although this cost is in general a rather complicated function involving line tariffs, computer equipment, and so on, the cost is basically determined by the center's size and location with respect to its community of users. An optimization problem thus arises to determine size and position of the I-R centers and an assignment of users to centers so as to minimize total cost. The techniques for optimization available to solve this problem do not guarantee to find a network whose cost is the absolute minimum. However they do generate a network whose cost is lowest among all networks which are in some sense "close"

to the "optimal" network.

For a given positioning of the set of I-R centers the optimal assignment of users to I-R centers is more-or-less straightforward. Each user or user-group is assigned to the center for which communication costs are minimum (usually the closest center). If such an assignment cannot be made due to center saturation, then one chooses the least expensive among the alternatives of replacing an existing assignment at the center, increasing the size of the center, or assigning the user group to another center. Finding the optimal positions of the centers is the more difficult part of the problem. The best available techniques for solution are iterative in nature. One begins with an arbitrary (but feasible) location of the centers. A computer program then computes cost for the configuration and makes perturbations in the center locations in an attempt to reduce the cost. Small "exploratory" perturbations may at first be made followed by larger perturbations in a direction that seems most likely to reduce cost. When further perturbations fail to reduce cost, the optimization is complete.

Once a location of the centers is determined, optimization techniques can be applied to determine network topology that minimizes intercenter communications costs. Such analyses have been carried out by Frank, et al in the design of the ARPA network.* Although intercenter cost optimization could be carried out within the loop that optimizes center location, these costs will hopefully be sufficiently insensitive of center location to be optimized separately outside of the loop that finds an optimal location for the centers.

---

*Frank, H., et al, "Topological Considerations in the Design of the ARPA Computer Network", Proceedings of the Spring Joint Computer Conference, 1970.

D. AUGMENTED-CATALOG INPUTTING

Staff Members

Mr. A. R. Benenfeld
Mr. L. E. Bergmann
Ms. M. A. Jackson
Ms. V. A. Miethe

Cataloger Assistants

Miss M. A. Flaherty
Miss L. A. Langille

Student Assistant

Mr. G. S. Tomlin

SUMMARY

As of 15 March 1972, 20,200 documents have been indexed and 19,750 documents have been completely processed into the data base. In order to conserve disk storage space, both the document-selection rate and the amount of information included in some of the earliest document records have been reduced.

A group of approximately 250 records relating to the activities of the Electrical Engineering Department during M.I.T.'s Independent Activities Period (IAP), January, 1972, has been added to the data base. The types of records and applications of specific fields for each record type are described.

PROCESSING OF INTREX DOCUMENTS

After a summer hiatus, a regular schedule of document addition to the Intrex data base has been resumed. Presently, the catalog is updated about every six weeks. As of 15 March 1972, approximately 20,200 documents were indexed, 20,150 catalog records were reviewed, 20,000 records were keyed, and 19,750 records were completely processed.

The Intrex program originally projected a data base of about 10,000 documents. This number has proven satisfactory for experimental purposes and, as the data base has grown toward twice the intended size, we have sought ways to conserve valuable disk storage space. One means for doing this, compaction of the data by encoding, has already been carried out (see 15 March 1970 Semi-annual Report). Two additional methods for limiting storage requirements have recently been instituted: the rate of selection of documents for the data base has been curtailed, and the information content of the earliest catalog records has been reduced.

Elimination of some journal titles as selection sources was considered the most effective means of limiting the document selection rate with little or no disruption to the timeliness or coverage of data base topics or to the selection routine itself. Of the previous 70 source journal titles, 39 source titles are being retained, based on criteria of user group interest, quality and topic coverage. The number of new data-base documents accepted per year is now estimated at 3400 articles, a 35-percent reduction from the previous annual selection rates of 5200 documents per year. The 20 source journals listed in Table IID-1 now provide approximately 90 percent of all the documents being indexed annually. For all practical purposes, this list is effective retroactive to January 1971.

Reduction of the information content of the earliest catalog records involved both the catalog record and the inverted files. Basic citation data, i.e., title, author, and journal location, as well as some small fixed-length

Table IID-1

List of Principal Current Intrex Source Journals

Journal Title	Estimated Number of Articles Selected per Year
Physical Review	480
Journal of Applied Physics	297
Metallurgical Transactions	280
Soviet Physics - Solid State	242
Physics Letters	220
Physical Review Letters	219
Physical Society of Japan Journal	193
Soviet Physics - JETP	153
Applied Physics Letters	128
Solid State Communications	122
Scripta Metallurgica	93
IEEE Proceedings	90
JETP Letters	87
Philosophical Magazine	84
Acta Metallurgica	81
IEEE Transactions on Magnetics	60
Applied Optics	58
Japanese Journal of Applied Physics	58
IEEE Transactions on Microwave Theory and Devices	56
Institute of Metals Journal	50

fields were retained in the catalog record. All other fields were omitted. The inverted files were condensed by eliminating all but the title-word references.

These steps saved about 80 percent of the storage space previously required for these catalog records. To date about 3300 of the earliest documents have been abbreviated in this manner.

IAP DATA BASE

As described in Section B, a group of approximately 250 records relating to the activities of the Electrical Engineering Department during M.I.T's January 1972 Independent Activities Period have been added to the data base. These records fall into two distinct categories: research-interest profiles of Electrical Engineering Department faculty members, and descriptions of specific minicourses, lectures, independent research, and other activities offered by members of the Electrical Engineering Department during IAP. A sample research-interest profile record is shown in Fig. IID-1, and an activity record is shown in Fig. IID-2.

Although the data for these records are not of a standard bibliographic nature, they meshed with the existing augmented-catalog structure with little difficulty because of the many parallels between the data elements. Application of individual fields differed slightly, depending on record type. The fields used for research-interest profile records are described below. In numerical order, they are:

Field 1 - Document Number: As with the computer-graphics records (see Semiannual Activity Report, 15 September 1971), these records were assigned a block of numbers so that the records would remain as a group, easily distinguishable from the regular Intrex data base.

Field 5 - Fiche: Since these records were complete in themselves — that is, the catalog record serves the function of full text — no microfiche copy (text access) was needed.

Field 21 - Author: The name of the faculty member whose interests were outlined was entered.

Field 22 - Affiliation: The faculty member's title (for example, "Associate Professor") was added to the standard entry, "M.I.T., Cambridge, Mass. Dept. of Electrical Engineering", when available.

Field 24 - Title: The uniform entry, "Research interests of M.I.T. faculty member — — " was used. (This uniform title entry provided inverted-file access to all research-interest profile records by typing: t research interest.)

```
  DOCUMENT 25755; Research interests of M.I.T. Faculty member
Arthur B. Baggeroer; Baggeroer, Arthur B.; M.I.T. Rm. 20A-209,
ext. 5287.
ONLINE                              (FIELD 4)
   12/17/71
AFFILIATION                         (FIELD 22)
   M.I.T., Cambridge, Mass.  Dept. of Electrical Engineering
RECEIPT                             (FIELD 46)
   05/00/71
ABSTRACT                            (FIELD 71)
   (See subject terms in field 73)
SUBJECTS                            (FIELD 73)
   E.E. Research Area 6 - Communication and Probabilistic Systems
   (0);
 E.E. Research Area 1 - Systems Science and Control Engineering
   (0);
 communications, state variable applications to communications
   (1);
 space-time and distributed random processes (1);
 sonar, underwater channels, and signal processing applications of
   oceanographic data (1);
```

Fig. IID-1 A Sample Research Interest Profile Record

```
  DOCUMENT 25903; Seminar on computer architecture; Madnick,
Stuart; Dates and time to be arranged. Orientation meeting - Wed.,
Jan. 5, 1972, 2 pm or contact instructor, Rm. 220 Tech Sq., ext.
5882.
ONLINE                              (FIELD 4)
   12/17/71
AFFILIATION                         (FIELD 22)
   :Instructor: M.I.T., Cambridge, Mass.  Dept. of Electrical
   Engineering. Rm. 220 Tech Sq., ext. 5882.
APPROACH                            (FIELD 66)
   Professional
ABSTRACT                            (FIELD 71)
   This seminar will discuss various types of computer systems,
   their differences, and reasons for differences.  Sample topics
   include: microprogramming, data organization, minicomputer
   design tradeoffs, reliability, organization for
   multiprogramming and timesharing, complex computer structures.
   Prereq.: 6.251 or 6.233 or 6.711 or 6.271 or 6.233; Enrollment
   limit : 15;
 Credit : 6 units available
SUBJECTS                            (FIELD 73)
   1972 Independent Activity Period (0);
 IAP Category - D-1 (0);
 E.E. Research Area 2 - Computer Science (0);
 seminar in architecture of computer systems (1);
 microprogramming, data organization, minicomputer design
   tradeoffs, reliability, multiprogramming and timesharing
   organization topics in computer systems (2);
 related subjects -  6.251, 6.233, 6.711, 6.271, 6.233 (4);
```

Fig. IID-2 A Sample IAP Activity Record

Field 46 - Article Receipt Date: This was a standard entry for all research-interest profile records. It indicated the date the information was originally published by the E.E. department in pamphlet form. It gives some indication of the currency of the information.

Field 47 - Location: Included the author's M.I.T. address and telephone extension. These data were included here so that they would be given automatically when normal output was requested.

Field 71 - Abstract: The abstract was replaced by a message referring the user to the subjects in Field 73 to avoid needless duplication of information.

Field 73 - Subjects:

(a) Subject phrases were derived directly from the Electrical Engineering Department publication, broken up into logical units, and each was assigned to range number 1, unless logical subordination was indicated. The subordinate terms were assigned to range number 2.

(b) As in the computer-graphics data base, the range-0 term was utilized as a quasi-classification technique. Here the classes were assigned according to seven major research groups in the Electrical Engineering Department. A separate range-0 term was created for each research group to which the faculty member's interests were related, as indicated in the source publication.

The application of fields for IAP-activity records differed from either the research-interest-profile records or the regular Intrex data base. Those which differ in application from the research-interest-profile records are described below.

Field 21 - Author: Includes the names of professors and other persons responsible for the specific activity.

Field 22 - Affiliation: Usually only the affiliation for the first faculty or staff member was entered. The M.I.T. address and telephone extension of the first faculty or staff member was included.

Field 24 - Title: Included the specific title of the activity preceded by a word or phrase describing the type of activity; for example, minicourse, lecture, research seminar, course extension, project laboratory, independent research.

Field 46 - Article Receipt Date: Not used in these records.

Field 47 - Location: There were four specific types of entries. These were used for: outlining actual meeting times and place for the activity; indicating that times and place were undecided and giving the name and address of the person to contact; indicating that the activity was an unscheduled one; and indicating the time and place of the orientation meeting. These types were used singly, or in combination, as the case required. The information was updated whenever necessary.

Field 65 - Author's Purpose and Field 66 - Level of Approach: These fields were utilized whenever applicable.

Field 71 - Abstract: An abstract was included for each activity. In addition, credit available, enrollment limit, and any prerequisites for the activity were described.

Field 73 - Subjects:

(a) Subject terms were derived from the description of the activity provided by the adviser. Range numbers were assigned considering the importance of the phrase in context.

(b) The range-0 term — "1972 Independent Activities Period" was added to each term set to provide a means of access to all IAP activity-type documents.

(c) An additional range-0 term, assigning each document to an activity type, i.e., lecture, mini-course, course extension, was created. The categories were indicated by single letters. This term was utilized in the publication of lists of activities under the various headings.

(d) Wherever possible, other range-0 terms indicating Electrical Engineering Research Groups to which the activity was most closely related were also included.

(e) Wherever applicable, a range-4 term listing M.I.T. subjects, by subject number, related to the specific activity was included.

E. COMPUTER SOFTWARE

Staff Members

Mr. C. E. Hurlburt
Mr. J. E. Kehr
Mr. M. K. Molnar
Dr. C. W. Therrien

SUMMARY

Changes have been made in the retrieval system to provide greater user control over primary searches, and to simplify the command language.

The data base has been updated to a total of 19,750 documents. Programs for partially removing catalog entries were completed and some 3300 references to early documents were reduced to contain only title, author, location, and a few short fields.

A special form of the Intrex programs was set up for retrieving information related to M.I.T's Independent Activities Period. Since information in this data base had to be current, a direct-editing facility for the data base was programmed.

The buffer/controller programs for the new BRISC console were placed in full operation. A few additional modifications were made as a result of user feedback regarding special features of the console.

INTREX RETRIEVAL SYSTEM SOFTWARE

Several changes and improvements have been made to the Intrex retrieval system. These changes were made both to simplify and extend features of the user interface.

Users may now control the stemming of search terms. Previously, all words typed in a subject-search request were stemmed. A search on reactors (stemmed to react) would also retrieve documents indexed under reactivation, reactivity, or reaction; a search statement containing the word past might retrieve documents indexed under paste or pastel, and so on. The new language feature permits users to override stemming by ending words with an exclamation point (!). Thus, the command

subject reactors!

retrieves only documents indexed under the unstemmed term reactors.

Another feature recently added to enhance user control of the search algorithm is word adjacency. A user may require that documents be retrieved only if their index phrases contain two specified search words immediately adjacent. He requests word adjacency by simply placing a hyphen (-) between the two search terms. For example the request

subject low-frequency response of germanium

would retrieve documents indexed under response of germanium diodes at low frequencies, but would not retrieve documents indexed under high frequency response of germanium at low temperatures.

The foregoing features serve to increase precision for users requiring a relatively narrow range of matches. Steps were taken to improve recall as follows. Previously a document would be retrieved only if all of the words in the subject search request were found within a single index phrase for the document. Now a document is retrieved if each of the words can be found anywhere within the set of index phrases for the document. This has the effect of changing the implied operation between search terms from the Intrex "with" to the Intrex "and". Of course, the explicit form of the "with" command is still available.

Other features added during the reporting period include a capability to refer to documents by their relative position on a list, a capability to rename lists, and "news" and "help" features. Through the news feature, users are advised at log-in time of system changes, new features, and special schedules of operation. The "help" command provides users with online up-to-date information for specific features (especially new features) of the language. The help command is a coordinated extension, not a replacement, of the long-existing Intrex "info" command.

Other changes in the user interface have been made which provide no new capability, but nevertheless improve clarity of the Intrex language. These include a restructuring of certain commands such as those used to save and retrieve lists of documents in a disk file. The most important modification,however, is the change to so-called immediately executable commands. Under this mode of operation, most multiple commands typed by the user on a single line are interpreted by the system as if the commands had been typed in sequence as separate requests. Previously, the language had a context sensitivity to commands typed on the same line and certain sequences produced ad hoc results that would not occur if the commands were issued separately. The new command syntax is in most respects functionally

identical to the older syntax and provides more flexibility. Other advantages are greater consistency and ease of learning.

The process of outputting catalog data to a user has been rendered more efficient through overlapped reading and writing. The system now reads into core document n+1 while it is printing catalog information for document n . With this improvement the computer time to output standard bibliographic information has been shortened by 40 percent.

The data base has been updated to 19,750 documents (see also Section II-D) and information about 3300 early documents has been reduced. The former document records have been reduced to title, author, location and a few short fields. They are retrievable only by means of index terms appearing in the title. Special modifications were made to the data-base-generation programs to accomplish the partial removal of these documents.

IAP RETRIEVAL SYSTEM

A special version of the Intrex retrieval programs was compiled in connection with Intrex's participation in the Independent Activities Period (IAP) at M.I.T. (see Section II-B). Since an entirely new data base consisting of ongoing activity descriptions was involved, the interactive dialog had to be appropriately modified. In addition, a simplified and more appropriate version of the online guide had to be created. At first, use of the IAP-Intrex system was restricted to a few consoles; later this restriction was eliminated so that IAP-Intrex could be used from any console on the M.I.T. campus.

In order to update data-base information such as activity schedules quickly and conveniently, a new editing facility was programmed in the form of an Intrex-system-programmer command. The facility permits replacement of the contents of most fields with information equal in length or shorter than the existing information. The editing feature should also have many uses for the regular Intrex bibliographic retrieval system — for example, in correcting misspellings or other errors in catalog records. This mode of correction is a much less expensive alternative to regeneration of the data base, even for the small IAP data base.

BUFFER/CONTROLLER SOFTWARE

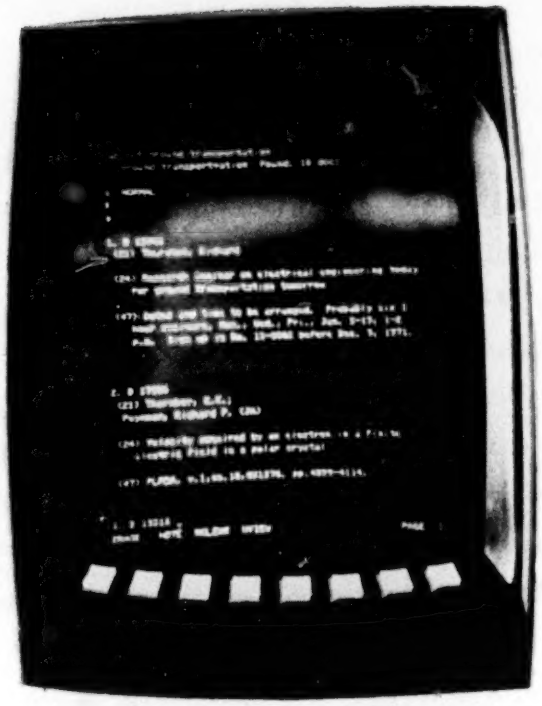
During this reporting period new buffer/controller software designed to permit concurrent use of two or more consoles was placed in operation. While reprogramming the buffer/controller, we took the opportunity to simplify and

improve console operation for the user. The two consoles currently in operation with the new software have been designated as BRISC, consoles (Bibliographic Remote Interactive Search Consoles).

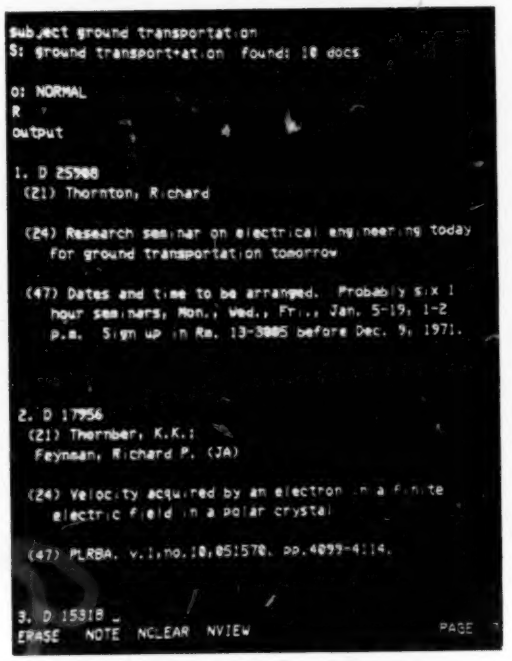
Recall that the BRISC console has several distinct modes of operation — the two primary modes being catalog mode and edit mode. Whenever a user is engaged in a dialog with the Intrex retrieval system, he is in catalog mode. During catalog mode up to 14 pages of the user's dialog are stored on the drum of the buffer/controller and may be selectively viewed by the user at any time. (We use the term "page" to mean a full screen of information on the cathode-ray tube.) A user may save pages or portions thereof by entering the edit mode and transferring catalog information that he desires to save to a special "note" page. Other modes of operation permit the user to view the special note page and to interact with the text access subsystem without also interacting with the catalog. During each mode of operation, functions related to turning pages, erasing pages, editing, switching modes, and so on are controlled by eight push-button function switches located beneath the screen. Since the functions performed by each button are programmed and change with the various modes of operation, labels for each button indicating its function in the current mode are displayed on the bottom line of the screen (see Fig. IIE-1).

The changes in the software extend features previously in existence on the BRISC and so render the console both more powerful and simple to use in carrying out a bibliographic search. One important change in catalog mode was the addition of a NEW PAGE function-switch. Previously a user conducting a search on the BRISC had two alternatives when the screen became full. First he could push the ERASE button to delete the information on the screen and then signal the central computer to continue interaction on the same page. Second, if he wished to save the screen information, he could press the NEXT PAGE button to select the next in the series of 14 available pages and then push ERASE to clear that page for interaction. With the addition of the NEW PAGE function, a user can both select the next page and clear it for interaction by pressing a single button. This feature greatly improves the rate at which a user can interact with the retrieval programs because it reduces the number of physical operations a user must perform and eliminates the delay encountered by the user's tendency to view the next page before erasing.

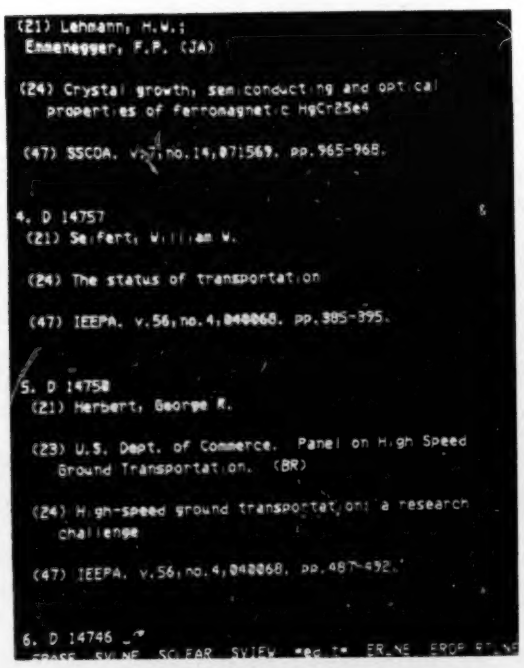
When the user has filled all 14 available pages and then requests a new page, the NEW PAGE function re-selects the first page but does not clear it for



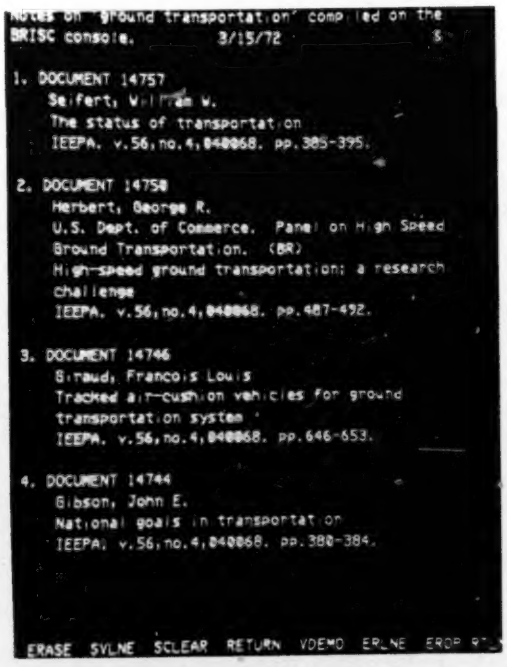
a) BRISC console screen showing programmed push-buttons.


subject ground transportation
S: ground transportation found: 10 docs
O: NORMAL
R
output
1. D 25988
(21) Thornton, Richard
(24) Research seminar on electrical engineering today for ground transportation tomorrow
(47) Dates and time to be arranged. Probably six 1 hour seminars, Mon., Wed., Fri., Jan. 5-19, 1-2 p.m. Sign up in Rm. 13-3085 before Dec. 9, 1971.
2. D 17956
(21) Thornber, K.K.;
Feynman, Richard P. (JA)
(24) Velocity acquired by an electron in a finite electric field in a polar crystal
(47) PLRBA. v.1,no.10,051570. pp.4099-4114.
3. D 15318
ERASE NOTE NCLEAR NVIEW
PAGE

b) Close-up of screen in Catalog mode.


(21) Lehmann, H.W.;
Emmenegger, F.P. (JA)
(24) Crystal growth, semiconducting and optical properties of ferromagnetic HgCr2Se4
(47) SSCOA. v.7,no.14,071569. pp.965-968.
4. D 14757
(21) Seifert, William W.
(24) The status of transportation
(47) IEEPA. v.56,no.4,040868. pp.385-395.
5. D 14758
(21) Herbert, George R.
(23) U.S. Dept. of Commerce. Panel on High Speed Ground Transportation. (BR)
(24) High-speed ground transportation; a research challenge
(47) IEEPA. v.56,no.4,040868. pp.487-492.
6. D 14746
ERASE SVLNE SCLEAR SVIEW

c) Screen in Edit mode.


Notes on "ground transportation" compiled on the BRISC console. 3/15/72
1. DOCUMENT 14757
Seifert, William W.
The status of transportation
IEEPA. v.56,no.4,040868. pp.385-395.
2. DOCUMENT 14758
Herbert, George R.
U.S. Dept. of Commerce. Panel on High Speed Ground Transportation. (BR)
High-speed ground transportation; a research challenge
IEEPA. v.56,no.4,040868. pp.487-492.
3. DOCUMENT 14746
Giraud, Francois Louis
Tracked air-cushion vehicles for ground transportation system
IEEPA. v.56,no.4,040868. pp.646-653.
4. DOCUMENT 14744
Gibson, John E.
National goals in transportation
IEEPA. v.56,no.4,040868. pp.380-384.
ERASE SVLNE SCLEAR RETURN VDEMO ERLNE EROP

d) Notes taken from Catalog output

Fig. IIE-1 BRISC console in various modes of operation showing labels for programmed buttons.

interaction. Instead, a message is flashed on the screen telling the user that he must push the ERASE button to continue interaction on this page. This procedure prevents inadvertent loss of information by automatic overwriting when the 14 pages are filled. The user may then erase the first page or he may use the NEXT PAGE button to select and view another page. If the new page has no information that he wishes to retain, then he may push ERASE to clear that page for interaction.

Since the pages reserved for catalog output are limited to 14, the user may want to condense some relevant bibliographic information and save it on the special "note" page. In order to simplify the compilation of this information certain functions previously available only in edit mode have been incorporated into the catalog mode of operation. One of these functions permits a user to transfer lines of information appearing in catalog mode to the note page for retention. The transfer is accomplished by positioning a special cursor at the line position and pressing the NOTE button (see Fig. IIE-1(b)). Two other buttons labeled NVIEW and NCLEAR permit the user to view the note page and to clear it of all existing information.

When it is desired to save lines of catalog information but to perhaps first modify or annotate them, the user enters edit mode. Here lines of catalog interaction can be modified by adding or deleting characters before they are transferred to the note page. Figure IIE-1(c) shows the console in edit mode, and Fig. IIE-1(d) shows the notes compiled from the catalog search on "ground transportation". The ampersand character (&) which appears near the top right of these photographs is the cursor.

F. HARDWARE

Staff Members

Mr. J. Bosco
Mr. P. Campoli
Mr. J. E. Kehr
Mr. D. R. Knudson
Professor J. F. Reintjes
Professor J. K. Roberge

SUMMARY

The second Intrex catalog-display console and its associated full-text display console have been installed in the Barker Engineering Library. The catalog console has been named BRISC (Buffered Remote Interactive Search Console) to distinguish it from other Intrex terminals. The principal problem in installing the BRISC concerned the two-way data communications via 1800 feet of coaxial cable connecting the console to the buffer/controller. A technique was developed for two-way transmission over a single coaxial cable.

An investigation of the page-centering problem on the text-access displays revealed that most of the positional error was caused by misalignment between the master and duplicate microforms during the production of duplicate fiche. Efforts to improve the process have resulted in significantly tighter positional tolerances. In addition to more accurate page centering, the scanning area has been reduced, thereby increasing the resolution of the text displays.

THE INTREX DISPLAY CONSOLES

The BRISC terminal in the Barker Engineering Library communicates with the buffer/controller via an 1800-foot length of coaxial cable. Digital data are transmitted over a single cable in both directions simultaneously; directional couplers are employed at each end to separate the incoming data from the combined signal.

High error rates were experienced during the initial tests because of wave-shape distortion resulting from the coaxial-cable frequency-response characteristics and the interference between incoming and outgoing signals at the terminal. Several minor design modifications were implemented to eliminate errors. The most significant improvement was achieved by reducing the transmission data rate from the console to the buffer/controller. A 1.2-megabit-per-second data rate

is needed from the buffer/controller to refresh the display, but the return data consists of keyboard inputs which can be accommodated by a substantially lower transmission rate. Tests showed that the interference between signals could be reduced to a tolerable level by reducing the console-to-buffer/controller transmission rate to 0.12 megabits per second. The current system has been operating for over three months with negligible transmission error rates.

The original BRISC console located in the Electronic Systems Laboratory (ESL) has been modified to permit its use as either an independent terminal or as a monitor-display slaved to the BRISC console in the Barker Library. In the monitor mode the terminal displays the same information that appears on the library terminal. The first four display lines can be used to enter instructional messages at the ESL terminal which are displayed on both terminals. The monitor display is useful for remote observation of user searches.

FULL-TEXT STORAGE AND RETRIEVAL

The centering of page images on full-text displays has been a long-standing problem. A closed-loop edge-finding technique using the flying-spot scanner has been employed to compensate for image positional inaccuracies in the horizontal direction of the page. However, there is no comparable technique employed for the vertical direction, with the result that the vertical position has been frequently unsatisfactory. An investigation of the various factors affecting positional tolerance showed that the most significant errors were caused by inaccurate page positioning on the microfiche itself. Misalignment of the stripped-up masters with the microfiche blanks caused poor page registration during generation of the fiche. Pages were required to fall within the grids of the COSATI specifications, but position variations from master to fiche could be a significant part of a page. Improved techniques in making fiche have allowed the vertical-position tolerance on the microfiche to be specified as ±0.010 inch of the nominal position. Existing fiche have been trimmed and reclipped where necessary to meet the tighter specifications. A series of repeatability tests made on the full-text retrieval equipment indicates that positional accuracy is now within 0.010 inch. The cumulative page-position error from all sources is less than ±3 percent of the vertical-page dimension.

The more accurate page registration not only improves page centering on full-text displays, but it also permits improvement in the system resolution. Previously, a 5.5-to-1 reduction ratio was used in the flying-spot scanner. This

ratio permitted the scanning raster to cover an area somewhat larger than the microfiche page image in an attempt to cover the complete page even if some positional misalignment occurred. With the tighter page-registration specifications, a 6-to-1 reduction ratio is employed, which reduces the effective scanner spot size and improves the system resolution.

III. MODEL LIBRARY PROJECT

A. STATUS OF THE PROJECT

Mr. C. H. Stevens
Mr. J. J. Gardner

The existing programs of the Model Library Project have continued and work has begun on several new programs. Response from institutions which have used Pathfinders and point-of-use programs remains encouraging and indicates the continued need for innovative solutions to traditional library problems.

The success of Pathfinders at M.I.T. and elsewhere reinforces the possibility that Pathfinder publication and distribution will evolve into a self-sustaining program. Efforts have been directed towards increasing editorial capability through involvement of subject librarians external to the Model Library program. Distribution of Pathfinders to other institutions has continued and feedback from these institutions will help in formulating policies concerned with the continuation and expansion of the program.

The point-of-use programs have been shared widely and are now in use in institutions in the United States, Canada, and Europe. Efforts have been made to increase the utility of the programs to other institutions by offering them in standard low-cost media, adaptable to inexpensive equipment. The enthusiastic response from users has led to work on additional programs.

Measurement of user preference of hard copy vs. microfiche copy continues to indicate preference for microfiche copy when it is provided at low cost for retention. The reduction of hard copy cost to users has had little effect on user preference. Maintenance of a relevant microform collection and availability of high quality reading equipment in a specially designed Microform Service Area plays an important role in this study.

Librarians from academic, public and special libraries attended five seminars, each of a full day's duration, as part of our visitor's program. Their response to our work has been encouraging and constructive. The visitor's program continues to provide a major means of increasing cooperation in the Pathfinder program and the sharing of the point-of-use programs. The advice and criticism of our visitors has been invaluable in directing our efforts.

New programs have been initiated during this reporting period. In response to an increase in substantive non-print research material the project staff will design a user-oriented non-print media area. The area will be designed for individual, on-demand use of a variety of media, including films and videotapes. Finally, as an adjunct to the point-of-use programs, work is under way on an audio-visual introduction to the Barker Engineering Library. Although designed for use in that particular library, the program will be adaptable to other engineering and science libraries.

B. POINT-OF-USE INSTRUCTION

Staff Members

Mr. C. H. Stevens
Mr. J. J. Gardner
Mrs. K. Boos
Miss M. P. Canfield

The six point-of-use instruction programs in operation in the Barker Engineering Library introduce users to the following reference sources: the author-title card catalog; the subject card catalog; Science Citation Index; Engineering Index; NASA STAR; and the Intrex augmented catalog and text-access systems. The catalog programs are synchronized sound-filmstrip; those on Engineering Index and the Intrex systems are synchronized sound-slide; and the two on NASA STAR and Science Citation Index are audio with sample pages. The sound-slide and audio with sample page programs are presented in units designed by the Model Library staff. The sound-filmstrip programs are presented in commercially available units.

The sound-slide units have proven to be superior to the sound-filmstrip equipment. Maintenance of the sound-slide units has been minimal, involving only occasional synchronization adjustment and lamp replacement. Editing of visuals is accomplished by deleting or replacing individual frames independently of the other frames in the program.

The sound-filmstrip units have a number of serious disadvantages. Each unit requires frequent cleaning of 12 lens and mirror surfaces; in-house production of filmstrips is more complicated and expensive than production of slides; and the editing of program visuals involves reshooting and reprocessing the entire filmstrip.

The audio with sample page units have some significant advantages over the audio-visual units. Because the audio track is the only program component, the programs and the presentation units are simpler and less expensive to produce. The chief advantage for the user is that with sample pages he sees abstracts or index entries in the context of an actual full-size page. This is not possible with slides or filmstrips.

New sound-slide equipment has been developed which will lower both production and display costs. More attractive and durable metal cabinets have been designed by the project staff and are being fabricated by a commercial metal

shop. The 1/4" audio tape cartridge player will be replaced in the new cabinets by a smaller cassette player utilizing continuous loop cassettes. Adoption of the cassette system will eliminate the necessity of using the facilities of a commercial recording studio for transferring the audio track from a master tape to a cartridge. The smaller, less expensive cassette player will account for a significant reduction in the size of the cabinet and will reduce the cost of the unit. Another important advantage is that the wide availability of cassettes facilitates the duplication and use of the programs by other institutions.

In addition to the new sound-slide unit, a new audio unit has been assembled at a cost of under $30. The unit consists of an inexpensive cassette player with rewind capabilities and a standard head set. It is currently in use with a program on Science Citation Index. The previously designed audio units are activated by simply lifting a phone, whereas the new units require the user to rewind, start and stop the program. However, the earlier units cost approximately $300 each. The major question is whether the substantial savings in cost justify using equipment which puts additional demands on the user. An obvious justification, however, is that $300 for one unit is a prohibitive cost for most libraries. This cost becomes a more important factor when installing a number of units covering a wide range of reference sources. User reaction to the new equipment is being measured through questionnaires.

Prior to this report, M.I.T.'s Audio-Visual Department was prepared to fabricate units for other institutions. An increased demand on that department's staff now makes this impossible. However, detailed plans and schematics for all equipment continue to be available on request so that institutions can assemble their own units with no design and development costs.

The instructional programs continue to be available in the form of tapes, slides and sample pages to outside institutions. Through a loan-duplication program, tapes and slides are mailed on request and returned after duplication by the borrowing institution. Institutions that have participated in the loan program include:

Bath University of Technology
University of California, Los Angeles (Biomedical Library)
University of California (San Francisco Medical Center)
University of Houston (Science Library)
University of Massachusetts (Amherst)

Oberlin College
Ohio State University
Rensselaer Polytechnic Institute
University of Southern Illinois
Stanford University
University of Western Ontario

The response from borrowing institutions has verified the need in research libraries for individualized instructional programs and inexpensive, simple ways to implement them. Efforts to enlarge the loan program continue, and feedback from other institutions will continue to influence our future program and equipment development.

The chief means of measuring user response to the programs in the Barker Engineering Library has been through the comment notebooks located with the displays. In order to gain a more objective measurement of user response a questionnaire was developed to supplement the comment notebooks. A sample of completed questionnaires large enough to provide reliable statistics is not yet available.

The comments to date on each program fall into categories as follows:

Science Citation Index	Responses	Percent
1. Uncritically favorable	5	56
2. Favorable with reservations	3	33
3. Favorable to the concept; unfavorable to specific program	1	11
4. Totally unfavorable	0	—
5. Irrelevant	0	—
6. Equipment problems	0	—

NASA STAR		
1. Uncritically favorable	24	56
2. Favorable with reservations	13	30
3. Favorable to the concept; unfavorable to specific program	0	0
4. Totally unfavorable	0	0
5. Irrelevant	4	9
6. Equipment problems	2	5

Engineering Index	Responses	Percent
1. Uncritically favorable	43	31
2. Favorable with reservations	44	32
3. Favorable to the concept; unfavorable to specific program	5	4
4. Totally unfavorable	7	5
5. Irrelevant	31	22
6. Equipment problems	9	6
Subject Card Catalog		
1. Uncritically favorable	27	31
2. Favorable with reservations	20	23
3. Favorable to the concept; unfavorable to specific program	14	16
4. Totally unfavorable	7	8
5. Irrelevant	15	17
6. Equipment problems	5	5
Author/Title Card Catalog		
1. Uncritically favorable	31	31
2. Favorable with reservations	33	33
3. Favorable to the concept; unfavorable to specific program	9	9
4. Totally unfavorable	5	5
5. Irrelevant	8	8
6. Equipment problems	14	14
Intrex Program		
1. Uncritically favorable	12	25
2. Favorable with reservations	7	14
3. Favorable to the concept; unfavorable to specific program	7	14
4. Totally unfavorable	1	2
5. Irrelevant	12	25
6. Equipment problems	10	20

The combined comments for all the programs fall into categories as follows:

Total Comments for All Programs	Responses	Percent
1. Uncritically favorable	137	32
2. Favorable with reservations	117	28
3. Favorable to the concept; unfavorable to specific program	35	8
4. Totally unfavorable	31	7
5. Irrelevant	70	16
6. Equipment problems	40	9

Two-thirds of the comments are favorable to the point-of-use concept while only seven percent are totally unfavorable. The favorable comments clearly indicate that users gained understanding of reference sources they previously knew little or nothing about. When judged advisable, recommendations made by users are incorporated into guidelines for revising existing programs and developing new programs.

An aspect of the programs which receives frequent comment is the use of irrelevant, light material inserted with the purpose of relaxing the listener. Most users indicate their approval of the use of this light material although some do not approve of the particular humor used. Only four percent of the comments indicated strong disapproval of the use of any humor.

Comments indicate a preference for audio with sample page programs over audio-visual programs. Comments of users who had used both include the following:

"Much clearer with notebook instead of slides."
"Very informative! I like notebook better than slides."
"Good — notebook better than slides."
"Seeing actual pages life-size is good."

Continuation of the point-of-use instruction program will emphasize new program production. Scripts for Chemical Abstracts, International Aerospace Abstracts, and Government Reports Index are in preparation. In addition the possibility of combining more than one reference source into a single program is under investigation. One such program is a combination of NASA STAR and

International Aerospace Abstracts, both of which cover the literature of aeronautics and the space sciences; NASA STAR, report literature and International Aerospace Abstracts, chiefly journal literature. Also under consideration is a combined program on Government Reports Index and the technical reports catalog in the Barker Engineering Library. The Government Reports Index serves as the subject index to report literature while the catalog functions as the holdings and location record for the library.

In addition to new program development, future activities will include final development and evaluation of new equipment; continued efforts to expand the loan-duplication program; and efforts to increase objective user feedback through questionnaires.

C. PATHFINDERS

Staff Members

Mr. C.H. Stevens
Mr. J.J. Gardner
Mrs. K.M. Boos
Miss M.P. Canfield
Mrs. R.J. Mead

Pathfinders currently available total 198: 147 in engineering and science; 34 in humanities; and 17 in social sciences. The titles are listed below in arbitrarily selected categories.

Architecture

Bauhaus
Byzantine Architecture
Medieval Architecture
Moorish Architecture
Romanesque Architecture in Europe

Art

American Folk Art
Book Illumination
Donatello
Encaustic Painting
Flemish Realism
French Impressionism
Giorgione
Monumental Brasses
Renaissance Art - Venetian School
Pierre-Auguste Renoir
J.M.W. Turner

Biomedical Engineering

Artificial Blood Circulation
Artificial Kidney
Artificial Limbs - Myoelectric Control
Artificial Organs - Heart
Blood - Circulation
Visual Perception

Business and Economics

Management Games
Minimum Wage Laws - U.S.

Civil Engineering

Air Conditioning
Airport Design
Asphalt
Coastal Engineering - Erosion
Earth Structures - Dams
Earthquake Engineering
Foundation Engineering
Ground Water Seepage
Harbor Design
Highway Engineering
Offshore Structures
Operations Research
Plates
Portland Cement Concrete
Rock Fracture/Failure
Salinity Intrusion
Sediment Transport
Shells
Soil Cements
Soil Freezing
Soil Instrumentation
Soil Stabilization
Systems Analysis
Thermal Stratification
Tunnels
Urban Mass Transportation

Civil Engineering (continued)

Urban Traffic Flow
Water Distribution Systems
Water Drainage Systems

Computer Technology

Analog Simulation
Analog-to-Digital Converters
Artificial Intelligence
Automata Theory
Cathode Ray Tube Display Devices
Cybernetics
Digital Simulation
Electronic Analog Computers
Electronic Digital Computers
Heuristic Programming
Holography
Hybrid Computers
Image Transmission Systems
Logic Design
Magnetic Disk/Drum Storage
Magnetic Tape Drive
Magnetic Tape Storage
Optical Character Recognition
Queuing Theory
Time Sharing

Education

Bilingual Education - U.S.
Education in Colonial New England
Sex Education

Electrical Engineering

Integrated Circuits
Lasers/Masers
Magnetohydrodynamics
Microwaves
Stroboscopic Photography
Telecommunication

Fluid Mechanics

Boundary Layer Control
Boundary Layer Flow
Boundary Layer Separation
Cavitation
Flow-Induced Vibrations
Fluidics

Fluid Mechanics (continued)

Laminar Boundary Layer
Noise Attenuation
Rheology
Thermal Boundary Layer
Turbulent Boundary Layer

Heat Transfer

Bubbles
Film Boiling
Heat Conduction
Heat Convection
Heat Transfer - Absorptivity
Heat Transfer - Emissivity
Nucleate Boiling
Plate-Fin Heat Exchangers
Pool Boiling
Radiation Heat Transfer
Shell-and-Tube Heat Exchangers
Thermal Contact Resistance
Thermal Regenerators
Thermal Stresses
Two Phase Flow

History

Great Proletarian Cultural Revolution
Medieval Manor
New South
U.S. Pacifism, 1940+

Information Science

Automatic Abstracting
Faceted Classification
Information Retrieval
Information Transmission
Scientific Information Transfer

Literature

Addison and Steele
Black American Novels - 20th Century
Classical Mythology in Modern English Literature
Eighteenth Century English Journalism

Literature (continued)

Etruscan Language
History of the English Bible
Icelandic and Old Norse Sagas
Jonathan Swift
La Pleiade
New England Transcendentalism
Puritanism in American Literature
Romantic Movement in German Literature

Materials Science and Engineering

Ferroelectrics
Ferromagnetism
Fiber Composite Materials
Mössbauer Effect
Particle Optics
Raman Effect
Superconductivity

Music

American Folk Music
Johann Sebastian Bach
Baroque Music
Early Music Printing
History of Opera
Igor Stravinsky

Ocean Engineering

Air Cushion Vehicles/Surface Effect Ships
Deep Sea Submergence Vehicles
Free Surface Hydrodynamics
Hydrofoils
Marine Power Systems
Marine Sonar Systems
Oil Pollution - Containment

Physics

Beam-Plasma Interactions
Time-Lag Systems
Underwater Acoustics

Political Science

Apportionment of State Legislatures
City Government - Council - Manager
Urbanization in America

Pollution

Air Pollution - Atmospheric Monitoring
Air Pollution - Automotive Exhaust Emissions
Air Pollution - Chimneys/Stacks
Air Pollution - Cyclones
Air Pollution - Electrostatic Precipitators
Air Pollution - Filtration
Air Pollution - Inversion Layers
Air Pollution - Nitrogen Oxides
Air Pollution - Plumes
Air Pollution - Radioactive Materials
Air Pollution - Scrubbers
Air Pollution - Smog
Air Pollution - Standards/Legislation-U.S.
Solid Waste Disposal - Composting
Thermal Pollution
Wastewater Treatment - Activated Sludge Process
Wastewater Treatment - Electrodialysis
Wastewater Treatment - Foam Fractionation
Wastewater Treatment - Sedimentation
Water Pollution - Detergents
Water Pollution - Mercury
Water Pollution - Monitoring Water Quality
Water Pollution - Phosphates
Water Pollution - Radioactive Materials

Science

Cartography
Crystal Defects
Glass
Group Theory
Insect Sex Attractants
Mathematical Logic
Neutrinos
Nucleation
Plant Physiology - Photosynthesis
Quarks
Set Theory
Soil Microbiology - Nitrogen Cycle

Science (continued)

Sunspots
Zeeman Effect

Sociology

Alienation
Juvenile Delinquency
Utopian Socialism
Witchcraft
Women's Liberation Movement in America

From this list of Pathfinders 86 (43%) were compiled by the Model Library staff. One hundred and twelve (57%) were supplied by participants in the cooperative program: 21 (11% of the total) were prepared by professional librarians; 91 (46% of the total) were compiled by library school students and edited by the Model Library staff. Clearly the library school students have been the most active contributors; however, the quality of their compilations has been uneven, resulting in an overall rejection rate of approximately 40%.

The student compilations that have been accepted for publication have required more editing time than had been anticipated. Because this time expenditure is unacceptable, the Model Library staff is attempting to effect a better screening process at the source through closer communication with the library school instructors who are the critical liaisons. The objective is to upgrade the quality of the student compilations forwarded to the project so that staff editing time will be reduced substantially. A major step in this direction has been taken by distributing to all cooperating library schools a revision of the Library Pathfinder Guidelines and Procedures which is shorter, more explicit, and more functionally oriented than the original version. In addition, the staff will allocate more time to systematic follow-up communication with instructors after their students' work has been received and reviewed. In this way the cooperating compilers will be given an appraisal of the general quality of the work submitted.

During calendar 1972, Pathfinders will be restricted to topics in current, high interest areas of engineering and the physical sciences. Compilers of social science and humanities Pathfinders who had previously participated in the cooperative program have been notified that the Model Library staff is no longer guaranteeing publication of non-science/technology compilations. However,

those library school instructors who judge Pathfinder compilation to be a meaningful and valid course requirement have been encouraged to continue the assignment and forward to us the work which they consider to have publication potential. These compilations will be held in reserve while negotiations with commercial publishers for Pathfinder distribution rights continue. These negotiations are discussed more fully in a subsequent part of this report.

The cost study on computer production of Pathfinders has been completed. The immediate expense of the computerized operation compared with manual production has been prohibitively high. Computer production affords long term advantages of format flexibility, ease of revision, and potentially, lower cost when large numbers of Pathfinders are to be handled. At present,however, only a small number of edited Pathfinder compilations are ready for production simultaneously and the manual system is adequate for the demands placed upon it. Because this situation is expected to continue during the coming year, computer production of Pathfinders is no longer under consideration.

During the period August through December, 1971, 574 Pathfinders were distributed on request to Barker Engineering Library users. Pathfinders on topics in the area of environmental pollution continue to account for the greatest number of requests. The acceptance of Pathfinders by librarians and users has been such that they are now considered an integral part of the Barker Engineering Library's augmented reference service.

Plans are under discussion by the staffs of the Model Library Project and the Barker Engineering Library to establish mechanisms within M.I.T. for wider and more frequent publicity of library services, including Pathfinders. This activity will be pursued in order to inform the large community of potential Pathfinder users — many of whom regularly use other libraries in the decentralized M.I.T. system — of the availability of these reference aids, especially a projected series on interdisciplinary topics.

During the last reporting period 55 Pathfinder compilations were received from 10 cooperating institutions. Twenty-two compilations were rejected, 5 social science compilations were removed from active consideration, and 28 science/technology Pathfinders are being considered for the edit-publication process (Fig. III-1).

Participants	Number of Pathfinder Compilations Submitted	General Subject Disciplines Represented
Case-Western Reserve University School of Library Science	1	Information Science
Clarkson College of Technology Burnap Memorial Library	1	Engineering
Engineering Societies Library	1	Engineering
St. Louis University Pius XII Memorial Library	1	Social Sciences
Simmons College School of Library Science	35	Science; Engineering
State University of New York at Buffalo Lockwood Memorial Library	1	Social Sciences
Texas Woman's University School of Library Science	8	Science
University of Houston Libraries Science Division	2	Science
University of Kentucky College of Library Science	1	Science
University of Pittsburgh Libraries	4	Social Sciences

Fig. III-1 Institutions Compiling Pathfinders in the Cooperative Program

Two institutions have elected to participate in the cooperative program on a trial basis by reviewing and editing library school student work. The Metropolitan Museum in New York is reviewing art compilations; the John D. Rockefeller, Jr. Library at Brown University is reviewing compilations on American literature topics.

Efforts are continuing to include more libraries in the cooperative program by involving them in systematically obtaining feedback from staff and patrons concerning the effectiveness of Pathfinders. The Model Library staff supplies Pathfinder masters and evaluative questionnaires to the cooperating libraries and makes procedural suggestions for handling the materials.

Student compilations have been given to a librarian at the Countway Library of Medicine for editing — not as part of the cooperative program but on a fixed fee per Pathfinder basis. It is anticipated that this involvement of a medical librarian as an editorial consultant will result in the preparation of a series of Pathfinders on medical topics of current interest for which the M.I.T. libraries do not have collection responsibilities.

With the cooperation of the Barker Engineering Library reference librarians, the Model Library staff has prepared a new reserve list of potential Pathfinder topics in current, high interest subject areas — primarily, air, water and noise pollution; solid waste disposal; waste reclamation and recycling; bioengineering; transportation; energy; materials science; and ocean engineering. The topics are being distributed to library school students on the basis of advance information from instructors concerning the user needs and subject collection strengths of the libraries in which the compilers will work. It is expected that professional librarians cooperating as compilers will continue to select topics based on their local collection strengths and their users' needs. All participants, however, will be requested to notify the Model Library staff of their final selections to avoid duplicate compilations.

Possibilities for the commercial publication of Pathfinders continue to be explored. The chief purposes of negotiations with commercial publishers are to expand this cooperative reference service and make it self-supporting. It has become increasingly obvious that many libraries would prefer to obtain Pathfinders by outright purchase rather than by committing staff time to compile, edit or evaluate them on a no-cost, cooperative basis. It is also clear that the Model Library staff does not have the staff capabilities required for marketing and distributing Pathfinders commercially.

Commercial publication would require that Pathfinders be compiled on topics in all disciplines in order to meet numerous and diverse reference needs. In addition to this quantitative expansion, it is likely that Pathfinders would be compiled for a wide range of library users including university, junior college and high school students. The responsibility for compiling and/or editing

Pathfinders of the established level of quality and preparing them to the stage of camera-ready copy would remain with the Model Library staff.

Discussions have been held with representatives of two major commercial publishers and preliminary proposals have been received from both. At present the two considerations under close study are the requirement for obtaining releases from compilers and the formulation of an agreement that will equitably satisfy the interests of Pathfinders users, the Model Library Project, the publishers, and the participants in the cooperative program.

D. USER PREFERENCE STUDY

Staff Members

Mr. J. J. Gardner
Miss C. L. Keator

During this reporting period there has been a significant increase in the number of users served by the Barker Engineering Library's Microform Service Area. The staff of the Barker Library has continued to emphasize the microform collection; reference librarians select professional society papers, engineering theses, government reports, and high demand journal and serial titles. The schedule of operations has been changed by the addition of evening hours until 9:00 p.m., Monday through Thursday.

Users of the Microform Service Area are offered a variety of microfiche readers including a vertical-screen microfiche viewer developed in the M.I.T. Electronic Systems Laboratory.* Indications are that the position control transport of this viewer is popular and that users judge image readability to be equal or superior to that of the portable and desk-top readers available. Equipment offered to library users is satisfactory to most, according to the results from questionnaires, shown in Fig. III-2.

	Number	Percent
Satisfied with equipment	214	82
Not satisfied	47	18
Total	261	100

Fig. III-2 Users' Evaluation of Microfiche Reading Equipment
7/1/71 - 1/1/72

Hard copy is made on a Xerox fiche-to-hard copy microprinter at a cost to the user of ten cents per page. Duplicate microfiche are made on a diazo process Bell and Howell Duplicator and given to users at no cost. During the eighteen-month period preceding July 1971, orders for microfiche accounted for 87 percent of the total orders processed. The latest six-month period showed an increase to 95 percent.

* For a description of this microfiche viewer, see Project Intrex Semiannual Activity Report 15 September 1971, pp. 95-97.

Type of Order	Number	Percent
Microfiche	477	95
Hard Copy	24	5
Total	501	100

Fig. III-3 Orders for Microfiche vs. Orders for Hard Copy
7/1/71 - 1/1/72

This increase appears to be the result of two factors: library users have become more aware of fiche storage advantages, and they are more comfortable in its use. Over the total two-year period, users have selected fiche over hard copy by a margin of 9 to 1, as listed in Fig. III-4.

Type of Order	Number	Percent
Microfiche	1150	90
Hard Copy	122	10
Total	1272	100

Fig. III-4 Orders for Microfiche vs. Orders for Hard Copy Cumulative Chart
1/1/70 - 1/1/72

Cost has not been the major factor in choosing fiche over hard copy during this reporting period. As indicated in Fig. III-5, it is more important to the user that fiche is compact and immediately available. During the period 1/1/70 - 6/30/71, 18 percent of the users requesting fiche did so out of curiosity; from 7/1/71 - 12/31/71, however, only 4 percent of the fiche requests were from users who wanted to "try it out". Of the 287 users choosing fiche during this reporting period, 49 percent of them did so because fiche was more convenient; this is an increase of 13 percent over the period 1/1/70 - 6/30/71.

Reason for Choosing Fiche	Number	Percent
1. Convenient	140	49
2. Immediately available	66	23
3. Less expensive	56	19
4. Curious about fiche	11	4
5. Miscellaneous	14	5
Total	287	100

Fig. III-5 Users' Reasons for Choosing Fiche over Hard Copy 7/1/71 - 1/1/72

A small number of users still prefer hard copy. Their reasons are shown in Fig. III-6.

Reasons for Choosing Hard Copy	Number	Percent
1. No reader available outside library	7	33
2. Dislike fiche	5	24
3. Need for frequent referral	3	14
4. Miscellaneous	6	29
Total	21	100

Fig. III-6 Users' Reasons for Choosing Hard Copy over Fiche 7/1/71 - 1/1/72

Despite the availability of portable loan readers, the most frequently cited reason for choosing hard copy is the lack of a fiche reader outside the library. A smaller number dislike fiche because they find it more difficult to read than hard copy.

The percentage of users choosing free fiche over hard copy at 10 cents per page remains approximately constant until document length exceeds 100 pages, after which all requests are for fiche. This range is shown in Fig. III-7.

Total Cost of Order at 10 Cents Per Page	Number Choosing Fiche	Number Choosing Hard Copy
$ .00 - .50	61 (91%)	6 (9%)
.51 .99	49 (97%)	2 (3%)
1.00 - 2.00	66 (96%)	3 (4%)
2.01 - 3.00	38 (86%)	6 (14%)
3.01 - 5.00	53 (95%)	3 (5%)
5.01 - 10.00	92 (96%)	4 (4%)
10.01 - 20.00	80 (100%)	0 (0%)
20.01 +	38 (100%)	0 (0%)

Fig. III-7 Correlation Between Cost of Hard Copy and Users' Choice of Fiche or Hard Copy 7/1/71 - 1/1/72

Fig. III-8 indicates the effect on user preference of a user reimbursement factor. There is an increase from 2 percent to 12 percent of users selecting hard copy when they are reimbursed by their department for the expense. The preference for fiche remains high, however, at 88 percent.

	Number Choosing Fiche	Number Choosing Hard Copy	Total
User would be reimbursed for expense	130 (88%)	17 (12%)	147
User would not be reimbursed for expense	212 (98%)	4 (2%)	216

Fig. III-8 Correlation Between Reimbursement of User and Choice of Hard Copy vs. Fiche 7/1/71 - 1/1/72

On December 6, 1971, the cost of hard copy was reduced to 5 cents per page. There has been no appreciable change in the percentage of requests for hard copy and fiche copy. This price will remain in effect for an experimental period, after which hard copy will be offered free in an effort to determine absolute preference between fiche and hard copy. An artificial time lag for the availability of fiche copies will be introduced to establish whether users will continue to prefer fiche if they must wait a time period equal to that which exists for hard copy. Additional models of fiche readers will be tested and features such as image sharpness, image brightness, image size, and ease of operation evaluated by users. Among these fiche readers will be some that can accommodate a 98-frame format.

This study clearly indicates that fiche are becoming more acceptable to library users and that, given the proper environment, users will choose fiche over hard copy. The environment should include fiche with good image quality, on-demand duplication service, and high quality fiche readers. It is important that portable fiche readers be available for loan, although an increasing number of offices and individuals are purchasing their own readers. It is also important that libraries be aware of literature available on fiche and maintain a current, relevant collection of material; the acceptance of fiche is at least related to its availability when full-size hard copy is unobtainable.

Studies in the Microform Service Area during the next reporting period will concentrate on equipment and types of fiche. Questionnaires will be introduced to gather data on user evaluations concerning various microfiche readers, the quality of duplicate fiche and hard copy, and desirable options in an ideal reader.

E. NEW PROGRAMS

Staff Members

Mr. J. J. Gardner
Mrs. K. M. Boos
Miss M. P. Canfield
Ms. R. J. Mead

Work on two projects has been initiated during this reporting period: installation and evaluation of a non-print media area; and production of a universally adaptable audio-visual introduction tc research engineering libraries.

NON-PRINT MEDIA AREA

A non-print media area is being developed to provide individual users with access to films and videotapes. Although there is an increasing amount of substantive research material in non-print form, many research libraries have not yet developed integrated media services. The Model Library will design an area in which the material itself will be readily available to the user and in which individualized projection equipment will be available for on-demand use. The media area will be designed to include an existing collection of 16 mm. sound films, 8 mm. cartridge films, and videotapes. All media represented in the collection will have projection equipment designed for individual viewing and listening. Users will have access to the materials and the projection equipment in one location.

Factors currently being considered are concerned with the physical location and arrangement of the area and equipment selection. The area must be isolated from the library study areas, yet easily accessible to the user. It must be comfortable and functional; adjustable lighting and sufficient power conduits are two elements receiving attention. Equipment must be easy to operate, inexpensive to maintain, and capable of being modified for individual screening and headset listening. User response to the area and its non-print material will be measured.

AUDIO-VISUAL ORIENTATION PROGRAM

The general orientation program is being scripted as a sound-slide program. The outline for the program is based on the Barker Library reference librarians' records of frequently asked questions. The program is in response to the recognition that the point-of-use programs function as a second level of orientation. The point-of-use programs on specific reference sources are effective only when

users are aware that the reference sources exist. This program will serve that function.

The program will be installed in a sound-slide unit similar to that used for point-of-use programs and will be available for individual use during all library hours. The evaluation process will consist of an informal comment notebook and a more formal questionnaire. The program will be available for loan and duplication by other institutions.

F. VISITOR'S PROGRAM

Staff Members

Mr. C. H. Stevens
Mr. J. J. Gardner
Miss M. P. Canfield
Mr. J. M. Kyed

Forty library administrators, representing American and Canadian academic, public and special libraries, participated in the formal visitor's program during this reporting period. The day-long programs included presentations on the Model Library project, the Barker Engineering Library and the Intrex augmented catalog and text-access experiments. The afternoon sessions of open discussion continued to be enthusiastic and constructive. Feedback worksheets which have been returned by participants indicate the visitor's program is a successful medium for describing new approaches to traditional library problems. In addition, many participants have joined our cooperative program for Pathfinder compilation and have borrowed and duplicated point-of-use programs for use in their own institution.

Visitors' suggestions on continuing and expanding the Pathfinder and point-of-use programs have been considered and work along the suggested lines has proceeded. It is of special concern to many visitors that the Pathfinder program be expanded in subject coverage and continued into the future. Representative comments also indicate interest in the expansion to other libraries of the Pathfinder cooperative program and the development and sharing of point-of-use hardware design.

The fruitful and rewarding exchange of ideas which has developed on each visitor's day makes continuation of the visitor's program a pleasant necessity. Programs will be held throughout 1972 and special programs are planned for the Special Libraries Association Annual Meeting, to be held this year in Boston.

IV. PROJECT INTREX STAFF

A. PROJECT OFFICE

Professor Carl F. J. Overhage, Director

Mr. Charles H. Stevens

B. ELECTRONIC SYSTEMS LABORATORY

Professor J. Francis Reintjes
Mr. Alan R. Benenfeld
Mr. Larry E. Bergmann
Mr. Joseph Bosco
Mr. D. J. Bottaro
Ms. Susan Foster Brown
Mr. Peter H. Campoli
Miss Margaret A. Flaherty
Mr. Charles E. Hurlburt
Ms. Margaret A. Jackson
Mr. Harold V. Jesse
Mr. James E. Kehr
Mr. Donald R. Knudson
Mr. Peter Kugel
Miss Linda A. Langille
Mr. Richard S. Marcus
Ms. Virginia A. Miethe
Mr. Michael K. Molnar
Professor James K. Roberge
Mr. James R. Sandison
Mr. F. Spahn
Dr. Charles W. Therrien
Mr. George S. Tomlin

C. BARKER ENGINEERING LIBRARY

Mr. James M. Kyed, Acting Head
Ms. Marjorie Chryssostomidis
Miss Barbara C. Darling
Ms. Kate Herzog
Miss Carol L. Keator
Miss Helen Magedson
Ms. Susan Nutter
Ms. Mary Pensyl
Ms. Carol Schildhauer
Mr. David C. Van Hoy

D. MODEL LIBRARY PROGRAM

Mr. Jeffrey J. Gardner
Mrs. Kathryn Boos
Miss Marie P. Canfield
Miss Molly Garfin
Ms. Renae Mead

V. CURRENT PUBLICATIONS

A. BOOK CHAPTERS, JOURNAL ARTICLES, AND CONFERENCE PAPERS

Benenfeld, A. R., and Marcus, R. S., "Intrex Subject Indexing and Its Relation to Classification". Presented at the American Society for Information Science Annual Meeting, Special Interest Group on Classification Research, Denver, Colorado, November 8, 1971.

B. INSTRUCTIONAL AIDS

Intrex Staff, "Reference Guide to Intrex", M.I.T. (Revised November 1971 — revision in press.)

Intrex Staff, "Summary Guide to IAP-Intrex", M.I.T., January, 1972.

VI. PAST PUBLICATIONS — October, 1969 through 15 September 1971

A. REPORTS

Hurlburt, C. E., Molnar, M. K., and Therrien, C. W., "The Intrex Retrieval System Software", ESL-R-458, September 15, 1971.

Uemura, S., "Intrex Subject/Title Inverted-File Characteristics", ESL-TM-454, September, 1971.

Goldschmidt, R. E., "File Design for Computer-Resident Library Catalogs", ESL-R-451, June, 1971. (Also a Ph.D. thesis, June 1971)

Goto, Nobuyuki, "A Translator Program for Displaying a Computer Stored Set of Special Characters", ESL-R-429, July, 1970.

Kusik, R. L., "A File Organization for the Intrex Information Retrieval System on the 360/67 CP/CMS Time-Sharing System". ESL-TM-415, January, 1970.

Lovins, J. B., "Error Evaluation for Stemming Algorithms as Clustering Algorithms", ESL-R-411, December, 1969.

Haring, D. R., "The Augmented-Catalog Console for Project Intrex (Part II)", ESL-TM-410, December, 1969.

Project Intrex Staff, Semiannual Activity Report, 15 September 1971.

Project Intrex Staff, Semiannual Activity Report, 15 March 1971.

Project Intrex Staff, Semiannual Activity Report, 15 September 1970.

Project Intrex Staff, Semiannual Activity Report, 15 March 1970.

B. BOOK CHAPTERS, JOURNAL ARTICLES, AND CONFERENCE PAPERS

Knudson, D. R., "An Experimental Text-Access System", to be presented at the XXIV Meeting of the Technical Information Panel of the Advisory Group for Aerospace Research and Development, NATO, September 9, 1971, Oslo, Norway.

Kugel, P., "Dirty Boole?" Journal of the American Society for Information Science, Vol. 22, No. 4, July, 1971, pp. 293-294.

Marcus, R. S., Benenfeld, A. R., and Kugel, P., "The User Interface for the Intrex Retrieval System". Presented at the Workshop on the User Interface for Interactive Search of Bibliographic Data Bases, Palo Alto, California, January 14-15, 1971. Proceedings to be published by AFIPS Press.

Lovins, J. B., "Error Evaluation for Stemming Algorithms as Clustering Algorithms". Journal of the American Society for Information Science, Vol. 22, No. 1, January, 1971, pp. 28-40.

Stevens, C. H., "Specialized Microform Applications in an Academic Library". Presented at the University of Denver, Denver, Colorado, December 7, 1970, at a Symposium on the Microform Utilization: The Academic Environment, 7-9 December, 1970, pp. 41-45.

Overhage, Carl F. J., "Directions for the Future", Presented at Collaborative Library Systems Development Conference, New York, N.Y., November 10, 1970. (Published in Conference Proceedings)

Reintjes, J. F., "Recent Experiments with the Project Intrex Information Storage and Retrieval System". Gordon Conferences, New London, New Hampshire, 16 July 1970.

Knudson, D. R., and Vezza, A., "Remote Computer Display Terminals". Conference on Computer-Handling of Graphical Information sponsored by SPSE, NMA, and SID, Newton, Mass., 9-10 July 1970, Proceedings, pp. 249-268.

Stevens, C. H., "New Whine in Olde Bottles", Presented at American Library Association National Convention, Detroit, Michigan, 2 July 1970.

Stevens, C. H., "Point-of-Use-Instruction in Libraries". Presented at American Library Association National Convention, Detroit, Michigan, 29 June 1970.

Stevens, C. H., "Destination Shangri-La, First Stop Erewhon". Presented at American Society for Engineering Education National Conference, Columbus, Ohio, 25 June 1970.

Roberge, J. K., and King, P. A., Jr., "An Economical Approach to High-Speed Character Generation and Display". 1970 Society for Information Display Symposium, New York, N.Y., 26-28 May 1970, Digest of Papers, pp. 104-105.

Stevens, C. H., "Experiments with Microfiche in an Academic Library". Presented at the National Microfilm Conference, San Francisco, California, 27 April 1970.

Reintjes, J. F., "Hardware", as related to "Issues and Problems in Designing a National Program of Library Automation". Library Trends, Vol. 18, No. 4, April, 1970, pp. 503-519.

Overhage, C. F. J., and Reintjes, J. F., "Computers in Libraries, Servant or Savant". Presented at American Society for Information Science, New England Chapter Meeting, 25 March 1970.

Knudson, D. R., "Image Storage and Transmission for Project Intrex". Conference on Image Storage and Transmission for Libraries, National Bureau of Standards, Gaithersburg, Maryland, 1-2 December 1969.

Overhage, C. F. J., "Information Networks", Chapter 11 in Annual Review of Information Science and Technology, Vol. 4, Carlos A. Cuadra, Editor. Encyclopedia Britannica, Inc., Chicago, 1969.

C. THESES

Chan, Y. T., "Full-Duplex Transmission of MHz Bipolar Digital Signals Over Coaxial-Cable Lengths Greater than 1,000 Ft.", Master of Science thesis, Electrical Engineering Department, Massachusetts Institute of Technology, June, 1971.

Goldschmidt, R. E., "File Design for Computer-Resident Library Catalogs", Ph.D. thesis, Electrical Engineering Department, Massachusetts Institute of Technology, June, 1971. (Also Electronic Systems Laboratory Report ESL-R-451.)

Goto, Nobuyuki, "A Translator Program for Displaying a Computer Stored Set of Special Characters". M.S. thesis, Electrical Engineering Department, Massachusetts Institute of Technology, July, 1970. (Also Electronic Systems Laboratory Report ESL-R-429.)

Kusik, R. L., "A File Organization for the Intrex Information Retrieval System on the 360/67 CP/CMS Time-Sharing System". M.S. thesis, Electrical Engineering Department, Massachusetts Institute of Technology, November, 1969. (Also Electronic Systems Laboratory Technical Memorandum ESL-TM-415.)

D. MISCELLANEOUS PRESENTATION

Charles H. Stevens

"The Role of Technology in Library Operation, Cooperation, and Architecture", Capital District Library Council, Schenectady, New York, August 17, 1971.

"Point-of-Use Instruction in Libraries", Greater Boston College and University Librarians, Waltham, Mass., June 10, 1971.

"Library Pathfinders", New England College Librarians Conference, Durham, N. H., April 17, 1971.

"A Model Approach to Library Instruction", Catholic Library Association, St. Louis, Missouri, March 27, 1971.

"Project Intrex and Engineering Library Services". Presented at Boston University, Boston, Massachusetts, 12 January 1971.

"Project Intrex at Midstream". Presented at the University of Illinois, Urbana, Illinois, 20 November 1970.

"The Sky is Not the Limit". Presented at Honeywell Corporation Executive Seminar, Concord, Massachusetts, 16 November 1970.

"Science and Technology Information Services in the Academic Library". Presented at North Carolina Central University, Durham, North Carolina, 21 October 1970.

* * * * *

Earlier publications and presentations are listed in previous issues of the Project Intrex Semiannual Activity Reports.